LEARNING BEYOND SITE BIAS FOR OOD GENERALIZATION IN BRAIN NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph-based learning from functional magnetic resonance imaging (fMRI) has shown strong potential for brain network analysis. However, existing methods often degrade under cross-site out-of-distribution (OOD) settings, as site-dependent confounder effects can obscure disease-related connectivity patterns and static functional connectivity (FC) does not explicitly capture informative within-scan variations. In this paper, we propose **C**ross-site **O**OD **R**obust brain n**E**twork (CORE), a unified framework for brain network learning across unseen sites. First, CORE estimates site-specific confounder effects and aggregates the resulting deconfounders by cross-source reliability for unseen-site correction, while extracting a population scaffold of reproducible label-associated connections from site-wise residualized FC. It then summarizes temporal variations on scaffold edges into compact descriptors for line-graph modeling. Finally, prior-guided subject-adaptive gating modulates message passing, balancing population priors with individual variability. Extensive leave-one-site-out experiments on ABIDE, REST-meta-MDD, SRPBS, and ABCD demonstrate that CORE consistently outperforms competitive baselines, with up to a 10.3% relative improvement in accuracy. These gains also persist across different brain parcellation schemes on ABIDE.

1 INTRODUCTION

Graph neural networks (GNNs) have shown strong potential for analyzing brain networks derived from functional magnetic resonance imaging (fMRI), with applications to autism spectrum disorder (ASD) and major depressive disorder (MDD) Bessadok et al. (2022); Luo et al. (2024). Conventional approaches estimate functional connectivity (FC) by quantifying statistical dependencies between blood oxygenation level-dependent (BOLD) time series from different regions of interest (ROIs) Finn et al. (2015); Kawahara et al. (2017). The resulting FC graphs enable GNNs to model functional relationships among brain regions and learn disease-related representations for computer-aided diagnosis Gu et al. (2025); Cui et al. (2022).

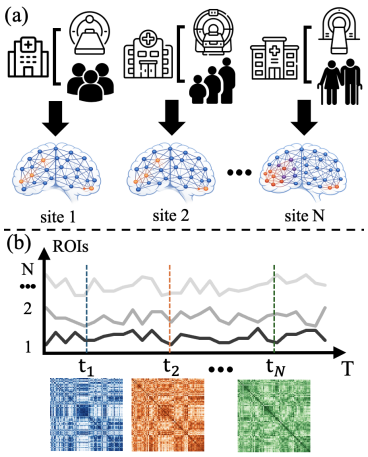


Figure 1: (a) Site-conditioned variations. (b) Time-varying FC.

Despite this progress, existing methods often degrade under cross-site out-of-distribution (OOD) settings Qiu et al. (2025; 2024). A key difficulty arises from structured, site-conditioned variations in FC associated with scanner characteristics and demographic composition Yamashita et al. (2025); Yu et al. (2018), as illustrated in Fig. 1(a). These effects can lead models to exploit site-related predictive shortcuts that may not remain reliable at unseen sites Brown et al. (2023); Chen et al. (2022); Xu et al. (2025). At the same time, static FC estimated over an entire scan does not explicitly capture within-scan connectivity variations that can carry complementary disease-related information Allen et al. (2014); Kaiser et al. (2016) (Fig. 1(b)). Although dynamic FC modeling incorporates these temporal variations, the underlying connectivity estimates remain sensitive to acquisition and sampling differences across sites Huotari et al. (2019); James et al. (2019), complicating the identification of transferable connectivity patterns.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

In this paper, we revisit cross-site OOD generalization for brain network analysis and aim to develop a unified framework. However, designing such a framework poses three key challenges: (1) *How can site-conditioned confounder bias be mitigated?* Pooled deconfounding with shared coefficients assumes site-invariant covariate effects, although the association between a given covariate and FC can differ across sites Xu et al. (2025); Qiu et al. (2025). For instance, when the association between age and a fronto-parietal FC edge varies across sites, subtracting a shared age effect can leave site-dependent residual bias and obscure reproducible disease-related connectivity patterns. (2) *How can dynamic FC be summarized compactly across scans of unequal length?* Dynamic FC is represented by high-dimensional connectivity sequences whose lengths and temporal resolution vary with scan duration and sampling protocols Chang et al. (2026); Huotari et al. (2019). These differences make it difficult to obtain comparable temporal representations across subjects and sites while retaining informative within-scan variations. (3) *How can cross-site regularities be reconciled with subject-specific variability?* Disease-related connectivity patterns that are reproducible across sites provide population-level guidance. However, FC varies across individuals Mueller et al. (2013); Yang et al. (2024), and the same patterns may not be equally informative for every subject. Models based solely on population patterns may overlook informative individual differences, whereas subject-specific modeling alone may retain site-related variations.

To tackle these challenges, we propose **C**ross-site **O**OD **R**obust brain n**E**twork (CORE), a unified framework with three components: (i) *Site-aware confounder decoupling* estimates edge-wise confounder effects per source site via Huber regression and aggregates deconfounders using cross-source reliability weights for unseen-site correction. From site-wise residualized FC, we extract a population scaffold of reproducible label-associated connections using consensus contrast magnitudes, directional consistency, and bootstrap stability. (ii) *Transient pathway profiling* summarizes sliding-window FC trajectories on scaffold edges into fixed-dimensional temporal descriptors, independent of scan length. These descriptors and deconfounded static FC form line-graph node features, with nodes representing scaffold connections linked through shared ROIs. (iii) *Prior-guided subject-adaptive gating* combines scaffold-derived population priors with subject-specific local pathway features and global context into adaptive gates that modulate line-graph message passing and readout, adapting pathway contributions to individual predictions. Leave-one-site-out (LOSO) evaluation on real-world benchmarks (ABIDE, REST-meta-MDD, SRPBS, and ABCD) shows that CORE outperforms competitive baselines, with up to a 10.3% relative gain in accuracy. Moreover, CORE remains robust to atlas variations, maintaining improvements across parcellation schemes.

Our contributions are summarized as follows: (1) We revisit cross-site OOD generalization in brain networks by jointly considering site-conditioned confounding, compact dynamic FC representations, and the balance between cross-site regularities and subject-specific variability. (2) We propose CORE, a unified framework that extracts a population scaffold from site-wise deconfounded FC to organize temporal pathway representations and guide subject-adaptive graph learning. (3) We conduct experiments on multiple real-world benchmarks, showing that CORE outperforms baseline methods under cross-site OOD settings and remains robust across different brain parcellation schemes.

2 Related Work

Graph Learning for Brain Network Analysis. Graph learning captures functional relationships among brain regions for subject-level prediction Li et al. (2021); Thapaliya et al. (2025); Gu et al. (2025). Static FC approaches estimate pairwise dependencies over full BOLD sequences, representing whole-brain connectivity without explicitly capturing within-scan variations. Dynamic FC and spatio-temporal graph models extend these representations by incorporating connectivity changes and temporal dependencies Hu et al. (2024); Kong et al. (2021). However, full dynamic FC trajectories are high-dimensional, and the underlying connectivity estimates are sensitive to acquisition and analysis choices, including sampling rate and windowing James et al. (2019); Preti et al. (2017); Huotari et al. (2019). Cross-site generalization further requires accounting for FC variations associated with acquisition protocols and population composition Qiu et al. (2025); Yamashita et al. (2025). In this context, CORE summarizes temporal variations on a population scaffold of reproducible label-associated connections into compact descriptors for subject-adaptive pathway modeling.

OOD Generalization on Graphs. Graph OOD generalization seeks robustness to distribution shifts through invariant representation learning, spurious subgraph pruning, and information bottleneck objectives Yao et al. (2024); Li et al. (2025a); Yao et al. (2025b;a). These strategies have been

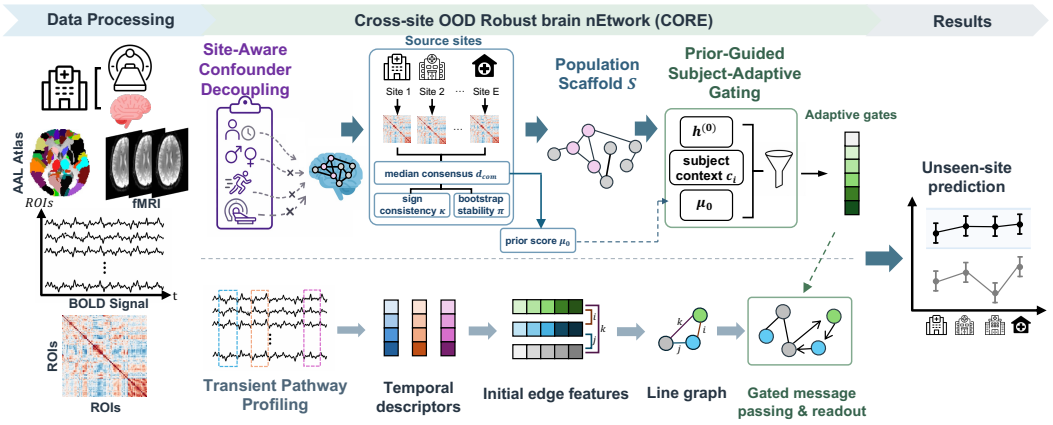


Figure 2: Overview of the proposed CORE. Site-Aware Confounder Decoupling mitigates site-conditioned bias and extracts a population scaffold of reproducible label-associated connections. Transient Pathway Profiling summarizes temporal variations on scaffold edges into lightweight descriptors and represents these connections as line-graph nodes. Prior-Guided Subject-Adaptive Gating combines scaffold priors with subject-specific features and context to modulate message passing for cross-site prediction.

explored in molecular and social graphs Zhang et al. (2026); Cai et al. (2025). Recent work on brain-network OOD learning combines feature filtering with consistent structure selection across subjects to identify predictive connections in static FC Xu et al. (2025); Zhao et al. (2026). Complementary neuroimaging research uses statistical harmonization to adjust site effects in FC measurements while accounting for observed covariates Yu et al. (2018). For unseen-site prediction, CORE estimates covariate effects separately at each source site and extracts a population scaffold of reproducible label-associated connections from residualized FC to guide subject-adaptive graph learning.

3 METHODOLOGY

Problem Setup. We study cross-site OOD generalization for brain network analysis, treating imaging sites as distinct environments. Let $\mathcal{E}_{\text{all}}$ denote all environments, with training environments $\mathcal{E}_{\text{tr}} \subset \mathcal{E}_{\text{all}}$ and unseen test environments $\mathcal{E}_{\text{te}} \subseteq \mathcal{E}_{\text{all}} \setminus \mathcal{E}_{\text{tr}}$. Each environment $e \in \mathcal{E}_{\text{all}}$ is associated with a distribution $\mathbb{P}^e$ over $(\boldsymbol{X}, \boldsymbol{q}, Y)$. For each $e \in \mathcal{E}_{\text{tr}}$, we observe a dataset $\widehat{\mathcal{D}}^e = \{(\boldsymbol{X}_i, \boldsymbol{q}_i, Y_i)\}_{i=1}^{N_e}$ of N_e subjects sampled from $\mathbb{P}^e$. Here, $\boldsymbol{X}_i \in \mathbb{R}^{T_i \times P}$ denotes the BOLD time series of subject i over P ROIs, with sequence length T_i varying across subjects. The vector $\boldsymbol{q}_i \in \mathbb{R}^d$ contains within-site covariates (age, sex, head motion) and site-level covariates (sequence length and TR), all available at inference. The binary label $Y_i \in \{0, 1\}$ indicates the control (0) or condition (1) group.

We consider predictors $f = \rho \circ T_{\widehat{\Theta}}$, where $T_{\widehat{\Theta}}(\boldsymbol{X}, \boldsymbol{q})$ maps BOLD time series and structured covariates to graph-level representations, $\widehat{\Theta}$ denotes the parameters and statistics estimated from training data, and ρ is the classifier. Our goal is to generalize from $\mathcal{E}_{\text{tr}}$ to unseen environments $\mathcal{E}_{\text{te}}$. We define the OOD risk as the average prediction risk across these environments:

$$\mathcal{R}_{\text{OOD}}(f) = \sum_{e \in \mathcal{E}_{\text{te}}} \frac{1}{|\mathcal{E}_{\text{te}}|} \mathbb{E}_{(\boldsymbol{X}, \boldsymbol{q}, Y) \sim \mathbb{P}^e} \left[l\big(f(\boldsymbol{X}, \boldsymbol{q}), Y\big) \right], \tag{1}$$

where $l(\cdot, \cdot)$ is the prediction loss.

Overview. We propose CORE, a unified framework for cross-site OOD generalization in brain network analysis (Fig. 2). It comprises three components: (i) **Site-Aware Confounder Decoupling** (Sec. 3.1) estimates site-specific deconfounders, aggregates them using cross-source reliability for unseen-site correction, and extracts a population scaffold of reproducible label-associated edges from site-wise residualized FC. (ii) **Transient Pathway Profiling** (Sec. 3.2) summarizes sliding-window FC trajectories on scaffold edges into fixed-dimensional temporal descriptors and represents these edges as line-graph nodes linked by shared ROIs. (iii) **Prior-Guided Subject-Adaptive Gating** (Sec. 3.3) combines scaffold-derived priors with subject-specific pathway features and context to modulate line-graph message passing and readout, producing subject-level representations.

3.1 SITE-AWARE CONFOUNDER DECOUPLING

Associations between observed covariates and FC can vary across imaging sites. We first characterize the residual bias that can arise when these associations are modeled using a shared linear deconfounding.

Let $\boldsymbol{r}_i \in \mathbb{R}^M$ denote the static FC vector of subject i from site e, obtained by stacking the upper-triangular entries (excluding the diagonal) of the full-scan ROI-wise Pearson correlation matrix in a fixed order, where $M = P(P-1)/2$. The edge index set is $\mathcal{I} = \{1, \dots, M\}$, with each $j \in \mathcal{I}$ corresponding to an undirected ROI pair (u_j, v_j) satisfying $1 \le u_j < v_j \le P$. We standardize the observed covariates as $\tilde{\boldsymbol{q}}_i = (\boldsymbol{q}_i - \boldsymbol{\mu}_q)/\boldsymbol{\sigma}_q$, where $\boldsymbol{\mu}_q, \boldsymbol{\sigma}_q \in \mathbb{R}^d$ are the covariate means and standard deviations estimated from pooled training subjects, and division is element-wise. A pooled linear deconfounding model computes

$$\hat{\boldsymbol{r}}_i^{\text{glob}} = \boldsymbol{r}_i - \boldsymbol{b}_{\text{glob}} - \boldsymbol{\Gamma}_{\text{glob}}^\top \tilde{\boldsymbol{q}}_i, \tag{2}$$

where $\boldsymbol{b}_{\text{glob}} \in \mathbb{R}^M$ and $\boldsymbol{\Gamma}_{\text{glob}} \in \mathbb{R}^{d \times M}$ are the edge-wise intercept vector and covariate coefficient matrix shared across sites.

Proposition 1 (Residual Bias of Pooled Deconfounding) *Assume FC follows the site-specific additive model*

$$\boldsymbol{r}_i = \boldsymbol{b}_e^\star + \boldsymbol{\Gamma}_e^{\star\top} \tilde{\boldsymbol{q}}_i + \boldsymbol{s}_i^\star + \boldsymbol{\epsilon}_i, \tag{3}$$

where $\boldsymbol{b}_e^\star \in \mathbb{R}^M$ and $\boldsymbol{\Gamma}_e^\star \in \mathbb{R}^{d \times M}$ are the site-specific intercept and covariate coefficients, $\boldsymbol{s}_i^\star \in \mathbb{R}^M$ is the confounder-free FC component with $\mathbb{E}[\boldsymbol{s}_i^\star \mid e, \tilde{\boldsymbol{q}}_i, Y_i] = \mathbb{E}[\boldsymbol{s}_i^\star \mid e, Y_i]$, and $\mathbb{E}[\boldsymbol{\epsilon}_i \mid e, \tilde{\boldsymbol{q}}_i, \boldsymbol{s}_i^\star, Y_i] = \mathbf{0}$. Let $\mathrm{D}_e[\boldsymbol{v}] = \mathbb{E}[\boldsymbol{v} \mid e, Y_i = 1] - \mathbb{E}[\boldsymbol{v} \mid e, Y_i = 0]$ denote the diagnostic contrast at site e. Then, for any fixed global coefficients $(\boldsymbol{b}_{\text{glob}}, \boldsymbol{\Gamma}_{\text{glob}})$,

$$\mathrm{D}_e\big[\hat{\boldsymbol{r}}_i^{\text{glob}}\big] = \mathrm{D}_e[\boldsymbol{s}_i^\star] + (\boldsymbol{\Gamma}_e^\star - \boldsymbol{\Gamma}_{\text{glob}})^\top \mathrm{D}_e[\tilde{\boldsymbol{q}}_i], \tag{4}$$

whereas the site-wise residual $\boldsymbol{r}_i - \boldsymbol{b}_e^\star - \boldsymbol{\Gamma}_e^{\star\top} \tilde{\boldsymbol{q}}_i$ satisfies $\mathrm{D}_e[\cdot] = \mathrm{D}_e[\boldsymbol{s}_i^\star]$.

Proposition 1 shows that pooled deconfounding contaminates the site-level diagnostic contrast whenever covariate effects differ across sites and the two groups differ in within-site covariates, so edges may be selected for reasons unrelated to diagnosis; site-wise deconfounding removes this term.

(i) Site-Aware Deconfounding. We estimate covariate-associated FC trends separately at each source site using Huber regression, which limits the influence of large residuals. For each training site $e \in \mathcal{E}_{\text{tr}}$ and edge $j \in \mathcal{I}$, we solve the following problem using iteratively reweighted least squares (IRLS) Fox & Weisberg (2002); Li & Swetits (1998); Holland & Welsch (1977):

$$(\hat{b}_{e,j}, \hat{\boldsymbol{\gamma}}_{e,j}, \hat{\theta}_{e,j}) = \arg\min_{b,\boldsymbol{\gamma},\theta} \sum_{i=1}^{N_e} \mathcal{H}_\delta\big(r_{i,j} - b - \tilde{\boldsymbol{q}}_i^\top \boldsymbol{\gamma} - \theta Y_i\big), \tag{5}$$

where $\hat{\theta}_{e,j}$ absorbs the diagnostic mean difference so that $(\hat{b}_{e,j}, \hat{\boldsymbol{\gamma}}_{e,j})$ estimates the covariate effect without removing the diagnostic contrast; only the covariate part is subtracted in Eq. (6), so no label is needed at inference. Stacking the estimates across edges gives $\hat{\boldsymbol{b}}_e = [\hat{b}_{e,1}, \dots, \hat{b}_{e,M}]^\top \in \mathbb{R}^M$ and $\hat{\boldsymbol{\Gamma}}_e = [\hat{\boldsymbol{\gamma}}_{e,1}, \dots, \hat{\boldsymbol{\gamma}}_{e,M}] \in \mathbb{R}^{d \times M}$. For scaffold estimation, we subtract the fitted site-specific trends from each source subject's FC:

$$\boldsymbol{r}_i^{\text{site}} = \boldsymbol{r}_i - \hat{\boldsymbol{b}}_e - \hat{\boldsymbol{\Gamma}}_e^\top \tilde{\boldsymbol{q}}_i, \qquad i = 1, \dots, N_e. \tag{6}$$

For unseen-site prediction, we aggregate source-site deconfounders according to their agreement on source covariates. Let $\mathcal{C}_e(\tilde{\boldsymbol{q}}) = \hat{\boldsymbol{b}}_e + \hat{\boldsymbol{\Gamma}}_e^\top \tilde{\boldsymbol{q}}$ denote the deconfounding map of site e. Its cross-source disagreement and normalized reliability weight are

$$u_e = \frac{1}{|\mathcal{E}_{\text{tr}}| - 1} \sum_{\substack{e' \in \mathcal{E}_{\text{tr}} \\ e' \neq e}} \frac{1}{N_{e'} M} \sum_{i=1}^{N_{e'}} \|\mathcal{C}_e(\tilde{\boldsymbol{q}}_i) - \mathcal{C}_{e'}(\tilde{\boldsymbol{q}}_i)\|_2^2, \qquad \alpha_e = \frac{(u_e + \varepsilon_u)^{-1}}{\sum_{k \in \mathcal{E}_{\text{tr}}} (u_k + \varepsilon_u)^{-1}}, \tag{7}$$

where the inner sum runs over subjects at site e' and $\varepsilon_u > 0$ ensures numerical stability. Smaller u_e yields a larger weight. We then form a frozen aggregate deconfounder and apply it to each unseen subject i using covariates standardized with the training statistics:

$$\bar{\boldsymbol{b}}_w = \sum_{e \in \mathcal{E}_{\text{tr}}} \alpha_e \hat{\boldsymbol{b}}_e, \qquad \bar{\boldsymbol{\Gamma}}_w = \sum_{e \in \mathcal{E}_{\text{tr}}} \alpha_e \hat{\boldsymbol{\Gamma}}_e, \qquad \boldsymbol{r}_i^{\text{res}} = \boldsymbol{r}_i - \bar{\boldsymbol{b}}_w - \bar{\boldsymbol{\Gamma}}_w^\top \tilde{\boldsymbol{q}}_i. \tag{8}$$

At an unseen site, this aggregate extrapolates source-site intercepts to its acquisition values and inherits a residual term as in Proposition 1, reduced but not removed by reliability weighting. Site-level covariates act only here, not in $\boldsymbol{r}_i^{\text{site}}$ or the scaffold. As residual FC is high-dimensional and unstable, we extract a population scaffold from site-wise residualized source data.

(ii) Cross-Site Scaffold Extraction. Using the site-wise residualized training FC $\boldsymbol{r}_i^{\text{site}}$, we extract a population scaffold $\mathcal{S} \subseteq \mathcal{I}$ of connections with reproducible label associations across source sites. For each source site e, let $\mathcal{D}_e^{(y)}$ index subjects with label $y \in \{0, 1\}$, with both groups represented. We define the site-level diagnostic contrast as $\boldsymbol{d}_e = \boldsymbol{\mu}_e^{(1)} - \boldsymbol{\mu}_e^{(0)}$, where $\boldsymbol{\mu}_e^{(1)}, \boldsymbol{\mu}_e^{(0)} \in \mathbb{R}^M$ are coordinate-wise Huber location estimates of $\boldsymbol{r}_i^{\text{site}}$ in the case and control groups, respectively.

To obtain a site-balanced consensus contrast, we aggregate $\{\boldsymbol{d}_e\}_{e \in \mathcal{E}_{\text{tr}}}$ using a fixed component-wise median rule:

$$\boldsymbol{d}_{\text{com}} \in \arg\min_{\boldsymbol{u} \in \mathbb{R}^M} \sum_{e \in \mathcal{E}_{\text{tr}}} \|\boldsymbol{d}_e - \boldsymbol{u}\|_1. \tag{9}$$

Each source site contributes equally, regardless of sample size. Huber location thresholds are fixed, and nonunique Huber location and median minimizers are resolved by midpoint selection. We assess directional consistency across source sites using κ_j and the stability of the consensus sign under site resampling using π_j:

$$\kappa_j = \frac{1}{|\mathcal{E}_{\text{tr}}|} \sum_{e \in \mathcal{E}_{\text{tr}}} \mathbb{I}[\text{sgn}(d_{e,j}) = \text{sgn}(d_{\text{com},j})], \quad \pi_j = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\Big[\text{sgn}(d_{\text{com},j}^{(b)}) = \text{sgn}(d_{\text{com},j})\Big]. \tag{10}$$

For each of the B bootstrap replicates, we sample $|\mathcal{E}_{\text{tr}}|$ source sites uniformly with replacement and recompute the component-wise median of the sampled site contrasts, retaining repeated draws. The resulting contrast for edge j is $d_{\text{com},j}^{(b)}$, and we define $\text{sgn}(0) = 0$. The population scaffold retains connections satisfying all three selection criteria:

$$\mathcal{S} = \{j \in \mathcal{I} \mid |d_{\text{com},j}| > \tau_E,\ \kappa_j \geq \eta_E,\ \pi_j \geq \zeta_E\}, \tag{11}$$

where τ_E, η_E, and ζ_E are selected on validation sites and all scaffold statistics are from training sites.

Proposition 2 (Margin Stability of the Site-Aware Scaffold) *For edge $j \in \mathcal{I}$, let $r_{i,j}^{\text{site}} = r_{i,j} - \hat{b}_{e,j} - \tilde{\boldsymbol{q}}_i^\top \hat{\boldsymbol{\gamma}}_{e,j}$ and $r_{i,j}^\star = r_{i,j} - b_{e,j}^\star - \tilde{\boldsymbol{q}}_i^\top \boldsymbol{\gamma}_{e,j}^\star$ denote the estimated and oracle site-wise residuals. Let $\mathcal{S}^\star$ be the scaffold obtained by applying the same selection rule, thresholds, and bootstrap draws to the oracle residuals, with oracle contrasts $d_{e,j}^\star$, $d_{\text{com},j}^\star$, and $d_{\text{com},j}^{\star(b)}$. Define the estimation perturbation and the oracle margin of edge j as*

$$\Delta_j = \max_{e \in \mathcal{E}_{\text{tr}}} \sum_{y \in \{0,1\}} \sup_{i \in \mathcal{D}_e^{(y)}} \left| r_{i,j}^{\text{site}} - r_{i,j}^\star \right|, \tag{12}$$

$$m_j^\star = \min\Big\{ \min_{e \in \mathcal{E}_{\text{tr}}} |d_{e,j}^\star|,\ \min_{1 \leq b \leq B} |d_{\text{com},j}^{\star(b)}|,\ |d_{\text{com},j}^\star|,\ \big||d_{\text{com},j}^\star| - \tau_E\big| \Big\}. \tag{13}$$

If $\Delta_j < m_j^\star$, then $j \in \mathcal{S}$ if and only if $j \in \mathcal{S}^\star$.

Proposition 2 shows that an edge retains its scaffold membership whenever the estimation error of the site-wise deconfounders is smaller than the smallest margin the selection rule depends on. Under the model of Proposition 1, this error is variance rather than bias and shrinks with site size, so edges with non-negligible oracle margins are selected consistently.

3.2 Transient Pathway Profiling

Static FC estimated over an entire scan does not explicitly capture within-scan connectivity variation. Full dynamic FC trajectories are high-dimensional and vary in length, complicating comparisons across subjects and sites Chang et al. (2026); Huotari et al. (2019). We therefore summarize temporal variations on the population scaffold $\mathcal{S}$ into fixed-dimensional descriptors that complement static FC.

Let $M_{\mathcal{S}} = |\mathcal{S}|$ and let $(j_p)_{p=1}^{M_{\mathcal{S}}}$ denote a fixed enumeration of the scaffold edges. For each subject i, we extract sliding windows of length W and stride S_w from $\boldsymbol{X}_i$, yielding $L_i = \lfloor (T_i - W)/S_w \rfloor + 1$ windows, where W and S_w are measured in time points. For each edge $j \in \mathcal{S}$ and window $t \in \{1, \ldots, L_i\}$, let $\boldsymbol{x}_{i,u_j}^{(t)}, \boldsymbol{x}_{i,v_j}^{(t)} \in \mathbb{R}^W$ denote the windowed signals from its two ROIs. We compute

the windowed Pearson correlation $\rho_{i,j}^{(t)} = \mathrm{corr}(\boldsymbol{x}_{i,u_j}^{(t)}, \boldsymbol{x}_{i,v_j}^{(t)})$ and apply the Fisher-z transformation $c_{i,j}^{(t)} = \frac{1}{2}\log\frac{1+\rho_{i,j}^{(t)}}{1-\rho_{i,j}^{(t)}}$.

We use temporal volatility $s_{i,j}$ to summarize dispersion across windows and temporal range $f_{i,j}$ to capture the full span of connectivity variation. Let $\bar{c}_{i,j} = \frac{1}{L_i}\sum_{t=1}^{L_i} c_{i,j}^{(t)}$ denote the temporal mean. The two descriptors are defined as

$$s_{i,j} = \sqrt{\frac{1}{L_i}\sum_{t=1}^{L_i}\left(c_{i,j}^{(t)} - \bar{c}_{i,j}\right)^2}, \quad f_{i,j} = \max_{1\le t\le L_i} c_{i,j}^{(t)} - \min_{1\le t\le L_i} c_{i,j}^{(t)}. \tag{14}$$

Both descriptors are invariant to time-constant additive shifts in the Fisher-z-transformed trajectories. Let $\boldsymbol{r}_i^{\mathrm{dec}}$ denote the deconfounded static FC, equal to $\boldsymbol{r}_i^{\mathrm{site}}$ for training-site subjects, whose own site offsets are removed exactly, and to $\boldsymbol{r}_i^{\mathrm{res}}$ otherwise, which removes the source-shared covariate effects but not the held-out site's own offset. For each scaffold edge j_p, we combine its deconfounded static FC with the log-transformed temporal descriptors to obtain the initial pathway representation:

$$\boldsymbol{h}_{i,p}^{(0)} = \left[r_{i,j_p}^{\mathrm{dec}}, \log(s_{i,j_p}+\varepsilon), \log(f_{i,j_p}+\varepsilon)\right]^\top \in \mathbb{R}^3, \tag{15}$$

where $\varepsilon > 0$ ensures numerical stability.

To model interactions among pathways sharing brain regions, we construct a shared line graph $\mathcal{G}_{\mathrm{line}} = (\mathcal{V}, \mathcal{E}_{\mathrm{line}}, \boldsymbol{A})$, where $\mathcal{V} = \{1, \ldots, M_{\mathcal{S}}\}$. Each node p represents scaffold edge j_p and carries subject-specific features $\boldsymbol{h}_{i,p}^{(0)}$. Distinct nodes p and p' are connected when their ROI pairs share an endpoint, i.e., $\{u_{j_p}, v_{j_p}\} \cap \{u_{j_{p'}}, v_{j_{p'}}\} \neq \emptyset$.

Line-graph edges receive larger weights when the corresponding pathways have similar absolute consensus contrasts. Let $a_p = |d_{\mathrm{com},j_p}|$ and $\boldsymbol{a} = [a_1, \ldots, a_{M_{\mathcal{S}}}]^\top$. We standardize these magnitudes to obtain population prior scores $\boldsymbol{\mu}_0 = (\boldsymbol{a} - \bar{a}\mathbf{1})/(\sigma_a + \varepsilon_\sigma)$, where

$$\bar{a} = \frac{1}{M_{\mathcal{S}}}\sum_{p=1}^{M_{\mathcal{S}}} a_p, \qquad \sigma_a = \sqrt{\frac{1}{M_{\mathcal{S}}}\sum_{p=1}^{M_{\mathcal{S}}}(a_p - \bar{a})^2}, \tag{16}$$

and $\varepsilon_\sigma > 0$ ensures numerical stability. The weighted adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{M_{\mathcal{S}} \times M_{\mathcal{S}}}$ is defined as

$$A_{pp'} = \begin{cases} \exp\left(-\dfrac{|\mu_{0,p} - \mu_{0,p'}|}{\Delta_{\mu_0} + \varepsilon_A}\right), & (p,p') \in \mathcal{E}_{\mathrm{line}}, \\ 0, & \text{otherwise}, \end{cases} \tag{17}$$

where $\Delta_{\mu_0} = \max_p \mu_{0,p} - \min_p \mu_{0,p}$ and $\varepsilon_A > 0$ prevents division by zero. Finally, we add self-loops to obtain $\hat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}$ and compute the symmetrically normalized propagation matrix $\tilde{\boldsymbol{A}} = \hat{\boldsymbol{D}}^{-1/2}\hat{\boldsymbol{A}}\hat{\boldsymbol{D}}^{-1/2}$, where $\hat{\boldsymbol{D}}$ is diagonal with $\hat{D}_{pp} = \sum_{p' \in \mathcal{V}} \hat{A}_{pp'}$.

3.3 Prior-Guided Subject-Adaptive Gating

The shared line graph provides a population-level structure for pathway modeling, while functional connectivity varies across individuals Mueller et al. (2013); Yang et al. (2024). The same pathways may not be equally informative for every subject. We therefore combine scaffold-derived priors with subject-specific features to adapt pathway contributions during message passing and readout.

Given the initial node features $\{\boldsymbol{h}_{i,p}^{(0)}\}_{p=1}^{M_{\mathcal{S}}}$, we compute a subject-level context by pooling the scaffold-node features:

$$\bar{\boldsymbol{h}}_i^{(0)} = \frac{1}{M_{\mathcal{S}}}\sum_{p=1}^{M_{\mathcal{S}}} \boldsymbol{h}_{i,p}^{(0)}, \qquad \boldsymbol{c}_i = \phi\left(\bar{\boldsymbol{h}}_i^{(0)}\right), \tag{18}$$

where $\phi(\cdot)$ is a multi-layer perceptron (MLP). For each node $p \in \mathcal{V}$, we combine its local features with the subject context and add the population prior score $\mu_{0,p}$ to the resulting logit:

$$\ell_{i,p} = \psi\left(\left[\boldsymbol{h}_{i,p}^{(0)} \,\|\, \boldsymbol{c}_i\right]\right) + \lambda\mu_{0,p}, \qquad g_{i,p} = \sigma\left(\frac{\ell_{i,p}}{\tau}\right), \tag{19}$$

where $\psi(\cdot)$ is an MLP producing a scalar output, $\|$ denotes concatenation, and $\sigma(\cdot)$ is the sigmoid function. The parameters $\lambda \geq 0$ and $\tau > 0$ control the prior strength and gate temperature, respectively. The resulting gate $g_{i,p} \in (0, 1)$ combines population guidance with the subject's local and global connectivity features.

For each subject, the gates are computed from the initial features and reused across L message-passing layers. Each gate scales the messages sent by its corresponding pathway, allowing subjects to assign different pathway weights within the shared propagation structure. At layer l, node p is updated as

$$\boldsymbol{h}_{i,p}^{(l)} = \boldsymbol{h}_{i,p}^{(l-1)} + \mathcal{F}^{(l)}\left(\boldsymbol{h}_{i,p}^{(l-1)}, \sum_{p' \in \mathcal{V}} \tilde{A}_{pp'}\, g_{i,p'}\, \boldsymbol{h}_{i,p'}^{(l-1)}\right), \tag{20}$$

where $\mathcal{F}^{(l)}(\cdot)$ is a learnable update function. Finally, we use the same gates to aggregate the refined node representations into a subject-level representation:

$$\boldsymbol{z}_i = \frac{\sum_{p=1}^{M_{\mathcal{S}}} g_{i,p} \boldsymbol{h}_{i,p}^{(L)}}{\sum_{p=1}^{M_{\mathcal{S}}} g_{i,p} + \varepsilon_z}, \tag{21}$$

where $\varepsilon_z > 0$ ensures numerical stability.

3.4 Learning Objective

Given the subject-level representation $\boldsymbol{z}_i$, we obtain the classifier output $\hat{y}_i = \rho(\boldsymbol{z}_i)$ using an MLP classifier ρ. Let $\widehat{\mathcal{D}}_{\mathrm{tr}} = \bigcup_{e \in \mathcal{E}_{\mathrm{tr}}} \widehat{\mathcal{D}}^e$ denote the pooled empirical training set. The classification loss is

$$\mathcal{L}_{\mathrm{cls}} = \mathbb{E}_{(\boldsymbol{X}_i, \boldsymbol{q}_i, Y_i) \sim \widehat{\mathcal{D}}_{\mathrm{tr}}}\big[\ell_{\mathrm{CE}}(\hat{y}_i, Y_i)\big], \tag{22}$$

where ℓ_{CE} denotes cross-entropy loss.

To control the total pathway weight assigned to each subject, we impose a soft gate-budget regularizer that penalizes deviations from a target gate mass:

$$\mathcal{L}_{\mathrm{budget}} = \mathbb{E}_{(\boldsymbol{X}_i, \boldsymbol{q}_i, Y_i) \sim \widehat{\mathcal{D}}_{\mathrm{tr}}}\left[\left|\sum_{p=1}^{M_{\mathcal{S}}} g_{i,p} - K\right|\right], \tag{23}$$

where $K \in (0, M_{\mathcal{S}}]$ is the target aggregate gate mass. The overall training objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{cls}} + \beta \mathcal{L}_{\mathrm{budget}}, \tag{24}$$

where $\beta \geq 0$ controls the gate-budget penalty.

4 Experiments

4.1 Experimental Settings

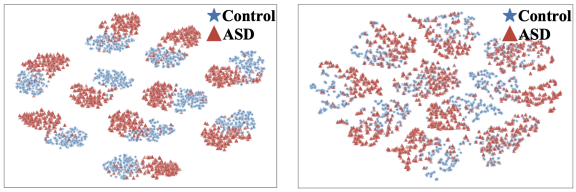
Datasets. To evaluate CORE under cross-site OOD settings, we use four real-world fMRI datasets spanning three disorders: ABIDE Di Martino et al. (2014) for ASD, REST-meta-MDD Yan et al. (2019) and SRPBS Tanaka et al. (2021) for MDD, and the large-scale developmental cohort ABCD Casey et al. (2018) for ADHD. These datasets exhibit substantial variation in demographics and acquisition protocols across sites and, for SRPBS, across acquisition batches. In addition to the standard ABIDE setting, we construct an ABIDE (CC200) variant using the CC200 functional parcellation to assess robustness to atlas variation. Further dataset details are provided in Appendix D.

Baselines. We compare the proposed CORE with a comprehensive set of baselines on the above datasets under the same cross-site OOD settings, spanning four categories: (1) general graph neural networks (GNNs), including GCN Kipf & Welling (2016), GAT Veličković et al. (2017), and GIN Xu et al. (2018); (2) general OOD methods, such as IRM Arjovsky et al. (2019) and CORAL Sun & Saenko (2016); (3) graph OOD methods, including GSAT Miao et al. (2022), DisC Fan et al. (2022), CEPG Wang et al. (2026), and DiSCO Sun et al. (2026); and (4) brain networks, including STAGIN Kim et al. (2021), BrainNetTF Kan et al. (2022), XG-GNN Qiu et al. (2024), AGMGC Noman et al. (2025), FC-HGNN Gu et al. (2025), LHDFormer Xue et al. (2025), NeuroH-TGL Li et al. (2025b), BrainOOD Xu et al. (2025), and DeCI Yu et al. (2026). More details of baselines are provided in Appendix E.

Table 1: Results on ABIDE, ABIDE (CC200), REST-meta-MDD, SRPBS, and ABCD (ADHD-task) over LOSO sites. **Bold** results indicate the best performance.

Type	Method	ABIDE		ABIDE (CC200)		REST-meta-MDD		SRPBS		ABCD (ADHD-task)	
		AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
GNN	GCN	62.13$_{\pm 10.32}$	58.64$_{\pm 9.66}$	50.43$_{\pm 8.86}$	50.41$_{\pm 8.55}$	60.64$_{\pm 4.18}$	56.32$_{\pm 2.87}$	73.81$_{\pm 19.46}$	70.23$_{\pm 19.17}$	71.75$_{\pm 6.52}$	63.14$_{\pm 9.15}$
	GAT	58.42$_{\pm 8.69}$	54.82$_{\pm 6.01}$	57.90$_{\pm 9.37}$	56.71$_{\pm 7.74}$	59.97$_{\pm 9.38}$	56.24$_{\pm 6.73}$	75.22$_{\pm 15.63}$	69.38$_{\pm 14.42}$	70.40$_{\pm 11.32}$	63.32$_{\pm 11.20}$
	GIN	57.95$_{\pm 5.21}$	54.60$_{\pm 3.10}$	52.61$_{\pm 8.83}$	49.86$_{\pm 10.23}$	56.58$_{\pm 6.72}$	53.47$_{\pm 4.68}$	72.33$_{\pm 21.67}$	68.74$_{\pm 18.09}$	68.97$_{\pm 8.84}$	60.73$_{\pm 7.03}$
OOD	CORAL	55.69$_{\pm 6.12}$	52.75$_{\pm 5.49}$	53.96$_{\pm 10.69}$	52.08$_{\pm 7.42}$	60.42$_{\pm 8.97}$	57.19$_{\pm 7.83}$	72.41$_{\pm 15.52}$	68.43$_{\pm 12.68}$	69.48$_{\pm 7.70}$	62.55$_{\pm 10.46}$
	IRM	58.77$_{\pm 6.12}$	55.50$_{\pm 3.85}$	51.98$_{\pm 10.64}$	52.86$_{\pm 6.99}$	58.63$_{\pm 4.38}$	54.22$_{\pm 3.41}$	69.48$_{\pm 21.87}$	67.82$_{\pm 16.78}$	70.40$_{\pm 6.60}$	59.06$_{\pm 7.57}$
Graph OOD	GSAT	58.69$_{\pm 7.62}$	56.12$_{\pm 7.98}$	51.02$_{\pm 7.05}$	51.07$_{\pm 6.27}$	57.38$_{\pm 5.82}$	54.83$_{\pm 4.47}$	71.32$_{\pm 18.96}$	67.68$_{\pm 15.83}$	71.27$_{\pm 8.73}$	61.16$_{\pm 7.48}$
	DisC	56.67$_{\pm 7.85}$	56.14$_{\pm 8.68}$	53.06$_{\pm 6.48}$	52.77$_{\pm 6.41}$	56.21$_{\pm 4.42}$	53.28$_{\pm 3.88}$	73.72$_{\pm 19.78}$	69.97$_{\pm 16.93}$	69.59$_{\pm 7.70}$	58.83$_{\pm 9.45}$
	CEPG	51.04$_{\pm 8.99}$	50.03$_{\pm 6.12}$	45.03$_{\pm 9.74}$	47.47$_{\pm 8.93}$	53.98$_{\pm 5.29}$	52.96$_{\pm 4.46}$	63.83$_{\pm 8.68}$	59.12$_{\pm 6.27}$	63.11$_{\pm 11.06}$	57.07$_{\pm 9.59}$
	DiSCO	56.78$_{\pm 6.72}$	55.47$_{\pm 4.83}$	51.58$_{\pm 8.45}$	51.37$_{\pm 6.21}$	57.96$_{\pm 3.73}$	54.08$_{\pm 2.84}$	74.18$_{\pm 18.18}$	67.64$_{\pm 15.46}$	67.51$_{\pm 9.83}$	61.98$_{\pm 7.44}$
Brain Networks	STAGIN	56.94$_{\pm 7.85}$	54.27$_{\pm 6.82}$	54.73$_{\pm 9.41}$	54.16$_{\pm 9.65}$	57.41$_{\pm 5.12}$	53.85$_{\pm 4.60}$	73.15$_{\pm 18.55}$	68.42$_{\pm 17.35}$	67.93$_{\pm 9.05}$	59.41$_{\pm 8.75}$
	BrainNetTF	62.85$_{\pm 6.20}$	60.90$_{\pm 7.06}$	57.04$_{\pm 11.21}$	52.55$_{\pm 8.08}$	61.50$_{\pm 7.40}$	58.20$_{\pm 5.60}$	61.56$_{\pm 17.03}$	60.41$_{\pm 10.80}$	68.28$_{\pm 4.54}$	56.91$_{\pm 10.43}$
	XG-GNN	60.41$_{\pm 7.73}$	54.12$_{\pm 4.09}$	50.64$_{\pm 12.57}$	52.71$_{\pm 8.78}$	60.08$_{\pm 7.23}$	56.24$_{\pm 6.49}$	81.80$_{\pm 15.21}$	74.69$_{\pm 17.29}$	71.20$_{\pm 5.46}$	55.66$_{\pm 9.78}$
	AGMGC	58.68$_{\pm 8.68}$	51.49$_{\pm 5.42}$	55.67$_{\pm 7.72}$	50.85$_{\pm 5.23}$	59.06$_{\pm 5.07}$	54.74$_{\pm 4.12}$	83.52$_{\pm 15.32}$	75.86$_{\pm 16.87}$	70.86$_{\pm 8.67}$	63.02$_{\pm 11.29}$
	FC-HGNN	55.26$_{\pm 7.90}$	52.23$_{\pm 4.58}$	55.04$_{\pm 9.71}$	53.21$_{\pm 5.45}$	58.69$_{\pm 6.69}$	56.78$_{\pm 5.71}$	57.30$_{\pm 10.60}$	53.80$_{\pm 6.40}$	71.08$_{\pm 5.74}$	59.10$_{\pm 12.34}$
	LHDFormer	57.63$_{\pm 7.61}$	55.39$_{\pm 6.74}$	56.45$_{\pm 9.11}$	55.27$_{\pm 9.48}$	59.84$_{\pm 5.88}$	55.43$_{\pm 4.95}$	75.63$_{\pm 17.20}$	69.87$_{\pm 16.05}$	69.52$_{\pm 8.44}$	61.28$_{\pm 9.15}$
	NeuroH-TGL	60.51$_{\pm 7.22}$	58.14$_{\pm 6.25}$	59.38$_{\pm 8.71}$	57.96$_{\pm 8.94}$	62.37$_{\pm 6.15}$	58.02$_{\pm 5.42}$	85.12$_{\pm 12.45}$	78.34$_{\pm 13.80}$	72.18$_{\pm 6.95}$	64.15$_{\pm 8.22}$
	BrainOOD	63.82$_{\pm 6.73}$	59.66$_{\pm 9.70}$	62.51$_{\pm 4.61}$	58.90$_{\pm 4.09}$	63.22$_{\pm 6.29}$	58.45$_{\pm 5.20}$	90.00$_{\pm 7.33}$	77.70$_{\pm 15.23}$	72.77$_{\pm 8.07}$	64.36$_{\pm 7.24}$
	DeCI	56.02$_{\pm 8.91}$	60.78$_{\pm 2.89}$	58.55$_{\pm 3.92}$	61.93$_{\pm 3.72}$	54.38$_{\pm 2.95}$	56.18$_{\pm 2.89}$	90.57$_{\pm 12.30}$	85.21$_{\pm 13.68}$	55.03$_{\pm 6.79}$	62.81$_{\pm 5.46}$
	CORE	**65.65**$_{\pm 6.42}$	**64.83**$_{\pm 6.62}$	**64.61**$_{\pm 7.12}$	**62.18**$_{\pm 8.35}$	**66.23**$_{\pm 9.35}$	**64.45**$_{\pm 5.15}$	**92.35**$_{\pm 7.04}$	**87.31**$_{\pm 10.52}$	**75.25**$_{\pm 7.63}$	**66.11**$_{\pm 8.65}$

4.2 Performance Comparison



(a) CORE (b) BrainOOD

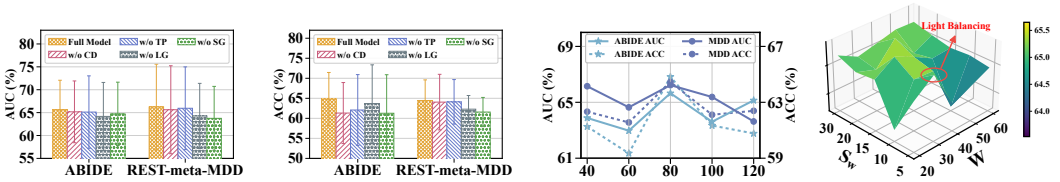
Figure 3: t-SNE visualizations on ABIDE.

We compare CORE with all baseline methods under the LOSO protocol on the ABIDE, REST-meta-MDD, SRPBS, and ABCD datasets in Table 1. We observe that: (1) Brain network methods achieve the best baseline results on all datasets and metrics, highlighting the importance of specialized FC modeling for cross-site neuroimaging prediction. (2) General OOD and graph OOD methods do not consistently improve performance across datasets, suggesting that generic robustness objectives alone may not adequately address site-conditioned variations in brain connectivity. (3) CORE consistently outperforms all baselines across datasets and metrics. These improvements reflect three complementary factors: (i) site-aware confounder decoupling adjusts site-dependent covariate effects, while population scaffold extraction retains reproducible label-associated connections; (ii) transient pathway profiling summarizes scaffold-level dynamic FC into compact, fixed-dimensional descriptors, providing temporal features that complement static connectivity; and (iii) prior-guided subject-adaptive gating combines population-level scaffold priors with individual pathway features and context to emphasize informative pathways during message passing and readout, adapting their contributions to each subject's prediction. Additionally, we present t-SNE visualizations on the ABIDE dataset under cross-site OOD settings for a strong baseline, BrainOOD Xu et al. (2025), and CORE, as shown in Fig. 3. The embeddings produced by CORE exhibit clearer separation between Control and ASD samples, indicating that it learns more discriminative representations under cross-site OOD settings.

To further assess whether the performance gains depend on a specific parcellation, we evaluate an additional ABIDE (CC200) setting by replacing the AAL atlas with the CC200 functional parcellation Craddock et al. (2012). CC200 comprises 200 functionally defined ROIs, resulting in different graph sizes, node definitions, and edge topologies compared to the anatomically defined AAL atlas. As shown in Table 1, CORE still outperforms baseline methods under this alternative parcellation, indicating that its gains are not solely dependent on the AAL atlas.

4.3 Ablation Study

To examine the contribution of each component in CORE, we conduct ablation studies on four variants: (1) CORE w/o CD, where site-aware deconfounding is replaced with global deconfounding; (2) CORE w/o TP, where transient pathway profiling is replaced by static FC features; (3) CORE w/o LG, where line-graph message passing is removed and scaffold nodes are independently encoded before gated readout; and (4) CORE w/o SG, where the gating module is removed and all scaffold pathways receive uniform weights during message passing and readout. As shown in Fig. 4(a, b), we observe that: (1) Replacing site-aware deconfounding with global deconfounding reduces performance, indicating the importance of accounting for site-dependent confounder effects. Removing TP also degrades



(a) AUC (b) ACC (c) Gate Budget K (d) Windowing (W, S_w)

Figure 4: Ablation results (a,b) and gate-budget sensitivity (c) on ABIDE and REST-meta-MDD, and windowing sensitivity (d) on ABIDE.

performance, suggesting that scaffold-level transient descriptors provide complementary dynamic information beyond static FC. (2) Removing LG or SG results in performance drops, highlighting the importance of interactions among scaffold pathways and subject-adaptive pathway selection for cross-site prediction. More results are provided in Appendix I.2.

4.4 Sensitivity Study

To assess how pathway weighting and temporal resolution affect cross-site prediction, we examine the sensitivity of CORE to the gate budget K, window length W, and stride S_w. As shown in Fig. 4(c), CORE achieves the highest AUC and ACC at $K = 80$ on both ABIDE and REST-meta-MDD, with lower performance at the other tested budgets. This pattern suggests a trade-off in pathway weighting: a small budget may overly suppress informative pathways, whereas a large budget may increase the contributions of less informative pathways during message passing and readout. We evaluate different combinations of W and S_w on ABIDE. As shown in Fig. 4(d), CORE achieves the best performance at $(W, S_w) = (30, 5)$. Larger window lengths or strides generally degrade performance. This suggests that longer windows may smooth out brief connectivity changes, while larger strides sample them less densely, limiting the temporal information captured by the descriptors. These results support moderate window lengths and fine strides for balancing temporal detail with stable FC estimation. More results are provided in Appendix I.3.

4.5 Interpretability Analysis

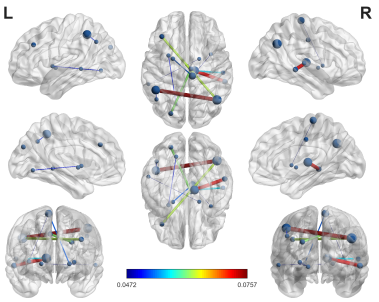


Figure 5: Top 10 positive scaffold connections on ABIDE.

To assess the neurobiological relevance of the scaffold connections, Fig. 5 visualizes the top ten positive scaffold connections on ABIDE under cross-site OOD evaluation. As these connections are selected from the population scaffold, they are label-associated edges reproducible across training sites. The resulting pattern is sparse and spans cortico-subcortical circuits, including parietal association regions, thalamocortical pathways, cingulo-insular nodes, and basal ganglia structures. These regions and circuits are broadly consistent with prior rs-fMRI findings in ASD, which implicate default-mode/fronto-parietal, thalamocortical, salience, and striatal/basal-ganglia connectivity Hull et al. (2017); Lee et al. (2016); Di Martino et al. (2011). Overall, these qualitative results suggest that the scaffold captures neurobiologically interpretable and cross-site reproducible connectivity patterns. A more detailed analysis of these positive connections is provided in Appendix I.4.

5 Conclusion

In this paper, we propose CORE, a unified framework for cross-site OOD generalization in brain network analysis. It combines site-aware deconfounding with reliability-weighted aggregation for unseen-site correction and uses a population scaffold to organize temporal pathway representations and guide subject-adaptive graph learning. Leave-one-site-out experiments on four multi-site datasets demonstrate consistent improvements in unseen-site prediction, with gains maintained under an alternative parcellation. Future work will explore acquisition-aware temporal descriptors and extend deconfounding to settings with incomplete covariate information.

AI USE STATEMENT

Large language models, including ChatGPT, were only used during manuscript revision for language polishing. The authors reviewed and revised the AI-assisted content and take full responsibility for the research ideas, experimental design, data analysis, and scientific conclusions.

REFERENCES

Elena A Allen, Eswar Damaraju, Sergey M Plis, Erik B Erhardt, Tom Eichele, and Vince D Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, 24(3):663–676, 2014.

Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019.

Alaa Bessadok, Mohamed Ali Mahjoub, and Islem Rekik. Graph neural networks in network neuroscience. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5833–5848, 2022.

Alexander Brown, Nenad Tomasev, Jan Freyberg, Yuan Liu, Alan Karthikesalingam, and Jessica Schrouff. Detecting shortcut learning for fair medical ai using shortcut testing. *Nature communications*, 14(1):4314, 2023.

Tingyi Cai, Yunliang Jiang, Yixin Liu, Ming Li, Changqin Huang, and Shirui Pan. Out-of-distribution detection on graphs: A survey. *arXiv preprint arXiv:2502.08105*, 2025.

Betty Jo Casey, Tariq Cannonier, May I Conley, Alexandra O Cohen, Deanna M Barch, Mary M Heitzeg, Mary E Soules, Theresa Teslovich, Danielle V Dellarco, Hugh Garavan, et al. The adolescent brain cognitive development (abcd) study: imaging acquisition across 21 sites. *Developmental cognitive neuroscience*, 32:43–54, 2018.

Xiao Chang, Tianye Jia, Zening Fu, Shitong Xiang, Yunman Xia, Chao Xie, Jiyan Zou, Miao Cao, Jie Zhang, Tobias Banaschewski, et al. Neurocognitive characterization of behaviour and mental illness through time-varying brain network analysis. *Nature Communications*, 2026.

Andrew A Chen, Joanne C Beer, Nicholas J Tustison, Philip A Cook, Russell T Shinohara, Haochang Shou, and Alzheimer's Disease Neuroimaging Initiative. Mitigating site effects in covariance for machine learning in neuroimaging data. *Human brain mapping*, 43(4):1179–1195, 2022.

R Cameron Craddock, G Andrew James, Paul E Holtzheimer III, Xiaoping P Hu, and Helen S Mayberg. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8):1914–1928, 2012.

Hejie Cui, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, and Carl Yang. Interpretable graph neural networks for connectome-based brain disorder analysis. In *International conference on medical image computing and computer-assisted intervention*, pp. 375–385. Springer, 2022.

Adriana Di Martino, Clare Kelly, Rebecca Grzadzinski, Xi-Nian Zuo, Maarten Mennes, Maria Angeles Mairena, Catherine Lord, F Xavier Castellanos, and Michael P Milham. Aberrant striatal functional connectivity in children with autism. *Biological psychiatry*, 69(9):847–856, 2011.

Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

Shaohua Fan, Xiao Wang, Yanhu Mo, Chuan Shi, and Jian Tang. Debiasing graph neural networks via learning disentangled causal substructure. *Proceedings of the Conference on Neural Information Processing Systems*, 35:24934–24946, 2022.

Emily S Finn, Xilin Shen, Dustin Scheinost, Monica D Rosenberg, Jessica Huang, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, 18(11):1664–1671, 2015.

John Fox and Sanford Weisberg. Robust regression. *An R and S-Plus companion to applied regression*, 91(6), 2002.

Yuheng Gu, Shoubo Peng, Yaqin Li, Linlin Gao, and Yihong Dong. Fc-hgnn: A heterogeneous graph neural network based on brain functional connectivity for mental disorder identification. *Information Fusion*, 113:102619, 2025.

Paul W Holland and Roy E Welsch. Robust regression using iteratively reweighted least-squares. *Communications in Statistics-theory and Methods*, 6(9):813–827, 1977.

Jinlong Hu, Jianmiao Luo, Ziyun Xu, Bin Liao, Shoubin Dong, Bo Peng, and Gangqiang Hou. Spatio-temporal learning and explaining for dynamic functional connectivity analysis: Application to depression. *Journal of Affective Disorders*, 364:266–273, 2024.

Jocelyn V Hull, Lisa B Dokovna, Zachary J Jacokes, Carinna M Torgerson, Andrei Irimia, and John Darrell Van Horn. Resting-state functional connectivity in autism spectrum disorders: a review. *Frontiers in psychiatry*, 7:205, 2017.

Niko Huotari, Lauri Raitamaa, Heta Helakari, Janne Kananen, Ville Raatikainen, Aleksi Rasila, Timo Tuovinen, Jussi Kantola, Viola Borchardt, Vesa J Kiviniemi, et al. Sampling rate effects on resting state fmri metrics. *Frontiers in neuroscience*, 13:279, 2019.

Oliver James, Hyunjin Park, and Seong-Gi Kim. Impact of sampling rate on statistical significance for single subject fmri connectivity analysis. *Human brain mapping*, 40(11):3321–3337, 2019.

Roselinde H Kaiser, Susan Whitfield-Gabrieli, Daniel G Dillon, Franziska Goer, Miranda Beltzer, Jared Minkel, Moria Smoski, Gabriel Dichter, and Diego A Pizzagalli. Dynamic resting-state functional connectivity in major depression. *Neuropsychopharmacology*, 41(7):1822–1830, 2016.

Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer. *Proceedings of the Conference on Neural Information Processing Systems*, 35:25586–25599, 2022.

Jeremy Kawahara, Colin J Brown, Steven P Miller, Brian G Booth, Vann Chau, Ruth E Grunau, Jill G Zwicker, and Ghassan Hamarneh. Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.

Byung-Hoon Kim, Jong Chul Ye, and Jae-Jin Kim. Learning dynamic graph representation of brain connectome with spatio-temporal attention. *Proceedings of the Conference on Neural Information Processing Systems*, 34:4314–4327, 2021.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Youyong Kong, Shuwen Gao, Yingying Yue, Zhenhua Hou, Huazhong Shu, Chunming Xie, Zhijun Zhang, and Yonggui Yuan. Spatio-temporal graph convolutional network for diagnosis and treatment response prediction of major depressive disorder from functional connectivity. *Human brain mapping*, 42(12):3922–3933, 2021.

Jung Min Lee, Sunghyun Kyeong, Eunjoo Kim, and Keun-Ah Cheon. Abnormalities of inter-and intra-hemispheric functional connectivity in autism spectrum disorders: a study using the autism brain imaging data exchange database. *Frontiers in neuroscience*, 10:191, 2016.

Haoyang Li, Xin Wang, Ziwei Zhang, and Wenwu Zhu. Out-of-distribution generalization on graphs: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025a.

Shengrong Li, Qi Zhu, Chunwei Tian, Xinyang Zhang, Wei Shao, Jie Wen, and Daoqiang Zhang. Neuroh-tgl: Neuro-heterogeneity guided temporal graph learning strategy for brain disease diagnosis. In *Advances in Neural Information Processing Systems*, volume 38, 2025b.

Wu Li and JJ Swetits. The linear l 1 estimator and the huber m-estimator. *SIAM Journal on Optimization*, 8(2):457–475, 1998.

Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H Staib, Pamela Ventola, and James S Duncan. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical image analysis*, 74:102233, 2021.

Xuexiong Luo, Jia Wu, Jian Yang, Shan Xue, Amin Beheshti, Quan Z Sheng, David McAlpine, Paul Sowman, Alexis Giral, and Philip S Yu. Graph neural networks for brain graph learning: A survey. *arXiv preprint arXiv:2406.02594*, 2024.

Siqi Miao, Mia Liu, and Pan Li. Interpretable and generalizable graph learning via stochastic attention mechanism. In *Proceedings of the International Conference on Machine Learning*, pp. 15524–15543. PMLR, 2022.

Sophia Mueller, Danhong Wang, Michael D Fox, BT Thomas Yeo, Jorge Sepulcre, Mert R Sabuncu, Rebecca Shafee, Jie Lu, and Hesheng Liu. Individual variability in functional connectivity architecture of the human brain. *Neuron*, 77(3):586–595, 2013.

Fuad Noman, Raphaël C-W Phan, Hernando Ombao, and Chee-Ming Ting. Adaptive graph learning with multi-graph convolutions for brain disorder classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 56–65. Springer, 2025.

Maria Giulia Preti, Thomas AW Bolton, and Dimitri Van De Ville. The dynamic functional connectome: State-of-the-art and perspectives. *Neuroimage*, 160:41–54, 2017.

Xinmei Qiu, Fan Wang, Yongheng Sun, Chunfeng Lian, and Jianhua Ma. Towards graph neural networks with domain-generalizable explainability for fmri-based brain disorder diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 454–464. Springer, 2024.

Xinmei Qiu, Yongheng Sun, Yilin Shi, Xujun Duan, Fan Wang, and Jianhua Ma. Metaexplainer: Revisit domain generalization of functional connectome analyses from the perspective of explainability. *Medical Image Analysis*, 105:103664, 2025.

Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *Proceedings of the European Conference on Computer Vision*, pp. 443–450. Springer, 2016.

Jerry Sun, Mohamed Abubakr Hassan, Yaoyu Zhang, Wanying Zhang, and Chi-Guhn Lee. Diverse and sparse mixture-of-experts for causal subgraph–based out-of-distribution graph learning. In *International Conference on Learning Representations*, volume 2026, pp. 79573–79597, 2026.

Saori C Tanaka, Ayumu Yamashita, Noriaki Yahata, Takashi Itahashi, Giuseppe Lisi, Takashi Yamada, Naho Ichikawa, Masahiro Takamura, Yujiro Yoshihara, Akira Kunimatsu, et al. A multi-site, multi-disorder resting-state magnetic resonance image database. *Scientific data*, 8(1):227, 2021.

Bishal Thapaliya, Esra Akbas, Jiayu Chen, Ram Sapkota, Bhaskar Ray, Pranav Suresh, Vince D Calhoun, and Jingyu Liu. Brain networks and intelligence: a graph neural network based approach to resting state fmri data. *Medical Image Analysis*, 101:103433, 2025.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Octave Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Shuo Wang, Mingchen Sun, Qiang Huang, and Ying Wang. Environment promoted invariant information learning for graph out-of-distribution generalization. *Artificial Intelligence*, pp. 104522, 2026.

Jiaxing Xu, Yongqiang Chen, Xia Dong, Mengcheng Lan, Tiancheng HUANG, Qingtian Bian, James Cheng, and Yiping Ke. BrainOOD: Out-of-distribution generalizable brain network analysis. In *Proceedings of the International Conference on Learning Representations*, 2025.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Rundong Xue, Xiangmin Han, Hao Hu, Zeyu Zhang, Shaoyi Du, and Yue Gao. Adaptive embedding for long-range high-order dependencies via time-varying transformer on fmri. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume 15971 of *Lecture Notes in Computer Science*, pp. 46–55. Springer Nature Switzerland, 2025.

Okito Yamashita, Ayumu Yamashita, Yuji Takahara, Yuki Sakai, Yasumasa Okamoto, Go Okada, Masahiro Takamura, Motoaki Nakamura, Takashi Itahashi, Takashi Hanakawa, et al. Computational mechanisms of neuroimaging biomarkers uncovered by multicenter resting-state fmri connectivity variation profile. *Molecular Psychiatry*, 30(11):5463–5474, 2025.

Chao-Gan Yan, Xiao Chen, Le Li, Francisco Xavier Castellanos, Tong-Jian Bai, Qi-Jing Bo, Jun Cao, Guan-Mao Chen, Ning-Xuan Chen, Wei Chen, et al. Reduced default mode network functional connectivity in patients with recurrent major depressive disorder. *Proceedings of the National Academy of Sciences*, 116(18):9078–9083, 2019.

Yanwu Yang, Chenfei Ye, Guinan Su, Ziyao Zhang, Zhikai Chang, Hairui Chen, Piu Chan, Yue Yu, and Ting Ma. Brainmass: Advancing brain network analysis for diagnosis with large-scale self-supervised learning. *IEEE transactions on medical imaging*, 43(11):4004–4016, 2024.

Tianjun Yao, Yongqiang Chen, Zhenhao Chen, Kai Hu, Zhiqiang Shen, and Kun Zhang. Empowering graph invariance learning with deep spurious infomax. *arXiv preprint arXiv:2407.11083*, 2024.

Tianjun Yao, Yongqiang Chen, Kai Hu, Tongliang Liu, Kun Zhang, and Zhiqiang Shen. Learning graph invariance by harnessing spuriosity. In *International Conference on Learning Representations*, volume 2025, pp. 18365–18401, 2025a.

Tianjun Yao, Haoxuan Li, Yongqiang Chen, Tongliang Liu, Le Song, Eric P. Xing, and Zhiqiang Shen. Pruning spurious subgraphs for graph out-of-distribution generalization. In *Proceedings of the Conference on Neural Information Processing Systems*, 2025b.

Guoqi Yu, Xiaowei Hu, Angelica I Aviles-Rivero, Anqi Qiu, and Shujun Wang. Moving beyond functional connectivity: Time-series modeling for fmri-based brain disorder classification. *IEEE Transactions on Medical Imaging*, 2026.

Meichen Yu, Kristin A Linn, Philip A Cook, Mary L Phillips, Melvin McInnis, Maurizio Fava, Madhukar H Trivedi, Myrna M Weissman, Russell T Shinohara, and Yvette I Sheline. Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fmri data. *Human brain mapping*, 39(11):4213–4227, 2018.

Kexin Zhang, Shuhan Liu, Song Wang, Weili Shi, Chen Chen, Pan Li, Sheng Li, Jundong Li, and Kaize Ding. A survey of deep graph learning under distribution shifts: from graph out-of-distribution generalization to adaptation. *ACM Transactions on Knowledge Discovery from Data*, 20(2):1–38, 2026.

Lin Zhao, Cheng Ding, Shengrong Li, Chunwei Tian, Chao Huang, Donghai Guan, Daoqiang Zhang, and Qi Zhu. A unified graph domain adaptation framework for cross-site brain network analysis. *IEEE Transactions on Medical Imaging*, 2026.

A NOTATION SUMMARY

Table 2: Key notation used in the main text.

Symbol	Meaning
$e \in \mathcal{E}_{\mathrm{tr}}, i$	Source-site and subject indices.
$(\boldsymbol{X}_i, \boldsymbol{q}_i, Y_i)$	BOLD time series, observed covariates, and binary label of subject i.
$j \in \mathcal{I}$	Canonical FC-edge index corresponding to ROI pair (u_j, v_j).
$\boldsymbol{r}_i, \tilde{\boldsymbol{q}}_i$	Full-scan static FC vector and covariates standardized using source data.
$\hat{\boldsymbol{b}}_e, \hat{\boldsymbol{\Gamma}}_e, \hat{\boldsymbol{\theta}}_e$	Estimated edge-wise intercepts, covariate coefficients, and label coefficients at source site e; only the first two are subtracted.
u_e, α_e	Cross-source disagreement score and reliability weight of deconfounder e.
$\bar{\boldsymbol{b}}_w, \bar{\boldsymbol{\Gamma}}_w$	Reliability-weighted aggregate coefficients for unseen-site correction.
$\boldsymbol{r}_i^{\mathrm{site}}, \boldsymbol{r}_i^{\mathrm{res}}$	Static FC after source-site correction and frozen aggregate correction, respectively.
$\boldsymbol{r}_i^{\mathrm{dec}}$	Static FC used for representation: $\boldsymbol{r}_i^{\mathrm{site}}$ for training-site subjects and $\boldsymbol{r}_i^{\mathrm{res}}$ otherwise.
$\boldsymbol{d}_e, \boldsymbol{d}_{\mathrm{com}}$	Site-level label contrasts and their component-wise median across source sites.
κ_j, π_j	Cross-site sign consistency and bootstrap sign stability of FC edge j.
$\tau_E, \eta_E, \zeta_E, B$	Contrast, sign-consistency, and bootstrap-stability thresholds for scaffold selection, and the number of bootstrap replicates.
$\mathcal{S}, M_{\mathcal{S}}$	Scaffold of reproducible label-associated FC edges and its size $M_{\mathcal{S}} = \lvert\mathcal{S}\rvert$.
p, p', j_p	Line-graph node indices p, p' and the original FC-edge index j_p corresponding to node p.
W, S_w, L_i	Sliding-window length, stride, and number of windows for subject i.
$c_{i,j}^{(t)}$	Fisher-z-transformed FC of edge j in window t for subject i.
$s_{i,j}, f_{i,j}$	Temporal volatility and range of the windowed Fisher-z FC trajectory.
$\boldsymbol{h}_{i,p}^{(0)}, \boldsymbol{h}_{i,p}^{(l)}$	Initial node features and layer-l representation of scaffold node p for subject i.
$\boldsymbol{\mu}_0$	Population prior scores obtained by standardizing absolute consensus contrasts on scaffold edges.
$\tilde{\boldsymbol{A}}$	Symmetrically normalized line-graph propagation matrix with self-loops.
$\boldsymbol{c}_i$	Subject context computed from pooled initial node features.
$g_{i,p}$	Continuous subject-adaptive gate for scaffold node p.
$\boldsymbol{z}_i$	Subject-level representation obtained through gated readout.
λ, τ, β	Prior strength and temperature of the gate, and weight of the gate-budget penalty.
K	Target total mass of the continuous gates.

As shown in the Table 2, we summarize the key notations of this paper.

B PROOF OF PROPOSITION 1

Proposition 1 (Residual Bias of Pooled Deconfounding). *Assume FC follows the site-specific additive model*

$$\boldsymbol{r}_i = \boldsymbol{b}_e^\star + \boldsymbol{\Gamma}_e^{\star\top} \tilde{\boldsymbol{q}}_i + \boldsymbol{s}_i^\star + \boldsymbol{\epsilon}_i, \tag{25}$$

where $\boldsymbol{b}_e^\star \in \mathbb{R}^M$ and $\boldsymbol{\Gamma}_e^\star \in \mathbb{R}^{d\times M}$ are the site-specific intercept and covariate coefficients, $\boldsymbol{s}_i^\star \in \mathbb{R}^M$ is the confounder-free FC component with $\mathbb{E}[\boldsymbol{s}_i^\star \mid e, \tilde{\boldsymbol{q}}_i, Y_i] = \mathbb{E}[\boldsymbol{s}_i^\star \mid e, Y_i]$, and $\mathbb{E}[\boldsymbol{\epsilon}_i \mid e, \tilde{\boldsymbol{q}}_i, \boldsymbol{s}_i^\star, Y_i] = \mathbf{0}$. Let $\mathrm{D}_e[\boldsymbol{v}] = \mathbb{E}[\boldsymbol{v} \mid e, Y_i = 1] - \mathbb{E}[\boldsymbol{v} \mid e, Y_i = 0]$ denote the diagnostic contrast at site e. Then, for any fixed global coefficients $(\boldsymbol{b}_{\mathrm{glob}}, \boldsymbol{\Gamma}_{\mathrm{glob}})$,

$$\mathrm{D}_e\big[\hat{\boldsymbol{r}}_i^{\mathrm{glob}}\big] = \mathrm{D}_e[\boldsymbol{s}_i^\star] + (\boldsymbol{\Gamma}_e^\star - \boldsymbol{\Gamma}_{\mathrm{glob}})^\top \mathrm{D}_e[\tilde{\boldsymbol{q}}_i], \tag{26}$$

whereas the site-wise residual $\boldsymbol{r}_i - \boldsymbol{b}_e^\star - \boldsymbol{\Gamma}_e^{\star\top}\tilde{\boldsymbol{q}}_i$ satisfies $\mathrm{D}_e[\cdot] = \mathrm{D}_e[\boldsymbol{s}_i^\star]$.

Proof.

Throughout, all random vectors have finite first moments, so that the conditional expectations below exist, and both diagnostic groups have positive probability at site e, so that $\mathrm{D}_e[\cdot]$ is well defined.

Since $\mathrm{D}_e[\cdot]$ is a difference of two conditional expectations, it inherits their linearity. For random vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^M$ with finite first moments,

$$\mathrm{D}_e[\boldsymbol{u} + \boldsymbol{v}] = \mathrm{D}_e[\boldsymbol{u}] + \mathrm{D}_e[\boldsymbol{v}]. \tag{27}$$

A fixed vector $\boldsymbol{c} \in \mathbb{R}^M$ has the same conditional expectation in both groups and therefore cancels in the difference,

$$\mathrm{D}_e[\boldsymbol{c}] = \mathbf{0}, \tag{28}$$

and for a fixed matrix $\boldsymbol{A} \in \mathbb{R}^{d\times M}$, linearity of conditional expectation gives

$$\mathrm{D}_e[\boldsymbol{A}^\top \tilde{\boldsymbol{q}}_i] = \boldsymbol{A}^\top \mathrm{D}_e[\tilde{\boldsymbol{q}}_i]. \tag{29}$$

The noise term also has zero contrast. By the tower property of conditional expectation and the assumption $\mathbb{E}[\boldsymbol{\epsilon}_i \mid e, \tilde{\boldsymbol{q}}_i, \boldsymbol{s}_i^\star, Y_i] = \mathbf{0}$, for each $y \in \{0, 1\}$,

$$\mathbb{E}[\boldsymbol{\epsilon}_i \mid e, Y_i = y] = \mathbb{E}\big[\, \mathbb{E}[\boldsymbol{\epsilon}_i \mid e, \tilde{\boldsymbol{q}}_i, \boldsymbol{s}_i^\star, Y_i] \,\big|\, e, Y_i = y\big] = \mathbf{0}, \tag{30}$$

and hence

$$\mathrm{D}_e[\boldsymbol{\epsilon}_i] = \mathbf{0}. \tag{31}$$

We now turn to the pooled residual. Fix a site e and global coefficients $(\boldsymbol{b}_{\mathrm{glob}}, \boldsymbol{\Gamma}_{\mathrm{glob}})$. Substituting the generative model into $\hat{\boldsymbol{r}}_i^{\mathrm{glob}} = \boldsymbol{r}_i - \boldsymbol{b}_{\mathrm{glob}} - \boldsymbol{\Gamma}_{\mathrm{glob}}^\top \tilde{\boldsymbol{q}}_i$ and collecting terms gives, for every subject i at site e,

$$\hat{\boldsymbol{r}}_i^{\mathrm{glob}} = \boldsymbol{s}_i^\star + \big(\boldsymbol{b}_e^\star - \boldsymbol{b}_{\mathrm{glob}}\big) + \big(\boldsymbol{\Gamma}_e^\star - \boldsymbol{\Gamma}_{\mathrm{glob}}\big)^\top \tilde{\boldsymbol{q}}_i + \boldsymbol{\epsilon}_i. \tag{32}$$

The four terms are the confounder-free component, an intercept mismatch that is constant across subjects at site e, a covariate-dependent slope mismatch, and noise. Applying $\mathrm{D}_e[\cdot]$ to both sides and expanding by Eq. (27),

$$\mathrm{D}_e\big[\hat{\boldsymbol{r}}_i^{\mathrm{glob}}\big] = \mathrm{D}_e[\boldsymbol{s}_i^\star] + \mathrm{D}_e\big[\boldsymbol{b}_e^\star - \boldsymbol{b}_{\mathrm{glob}}\big] + \mathrm{D}_e\big[(\boldsymbol{\Gamma}_e^\star - \boldsymbol{\Gamma}_{\mathrm{glob}})^\top \tilde{\boldsymbol{q}}_i\big] + \mathrm{D}_e[\boldsymbol{\epsilon}_i]. \tag{33}$$

The intercept term vanishes by Eq. (28), since $\boldsymbol{b}_e^\star - \boldsymbol{b}_{\mathrm{glob}}$ is fixed given e; the slope term equals $(\boldsymbol{\Gamma}_e^\star - \boldsymbol{\Gamma}_{\mathrm{glob}})^\top \mathrm{D}_e[\tilde{\boldsymbol{q}}_i]$ by Eq. (29) with $\boldsymbol{A} = \boldsymbol{\Gamma}_e^\star - \boldsymbol{\Gamma}_{\mathrm{glob}}$; and the noise term vanishes by Eq. (31). Substituting back into Eq. (33) yields

$$\mathrm{D}_e\big[\hat{\boldsymbol{r}}_i^{\mathrm{glob}}\big] = \mathrm{D}_e[\boldsymbol{s}_i^\star] + (\boldsymbol{\Gamma}_e^\star - \boldsymbol{\Gamma}_{\mathrm{glob}})^\top \mathrm{D}_e[\tilde{\boldsymbol{q}}_i], \tag{34}$$

which is the first claim of the proposition.

The site-wise residual is handled in the same way. Substituting the generative model gives

$$\boldsymbol{r}_i - \boldsymbol{b}_e^\star - \boldsymbol{\Gamma}_e^{\star\top} \tilde{\boldsymbol{q}}_i = \boldsymbol{s}_i^\star + \boldsymbol{\epsilon}_i, \tag{35}$$

so that, by Eqs. (27) and (31),

$$\mathrm{D}_e\big[\boldsymbol{r}_i - \boldsymbol{b}_e^\star - \boldsymbol{\Gamma}_e^{\star\top} \tilde{\boldsymbol{q}}_i\big] = \mathrm{D}_e[\boldsymbol{s}_i^\star] + \mathrm{D}_e[\boldsymbol{\epsilon}_i] = \mathrm{D}_e[\boldsymbol{s}_i^\star], \tag{36}$$

which is the second claim.

The decomposition in Eq. (32) also describes the bias at the subject level. Conditioning it on $(e, \tilde{\boldsymbol{q}}_i, \boldsymbol{s}_i^\star)$ rather than on (e, Y_i), and using $\mathbb{E}[\boldsymbol{\epsilon}_i \mid e, \tilde{\boldsymbol{q}}_i, \boldsymbol{s}_i^\star] = \mathbf{0}$, which again follows from the stated assumption by the tower property, gives

$$\mathbb{E}\big[\hat{\boldsymbol{r}}_i^{\mathrm{glob}} \mid e, \tilde{\boldsymbol{q}}_i, \boldsymbol{s}_i^\star\big] = \boldsymbol{s}_i^\star + (\boldsymbol{b}_e^\star - \boldsymbol{b}_{\mathrm{glob}}) + (\boldsymbol{\Gamma}_e^\star - \boldsymbol{\Gamma}_{\mathrm{glob}})^\top \tilde{\boldsymbol{q}}_i. \tag{37}$$

Both mismatch terms therefore bias the pooled residual relative to $\boldsymbol{s}_i^\star$ for each subject, but only the slope mismatch survives in the diagnostic contrast, where it enters through the between-group covariate difference $\mathrm{D}_e[\tilde{\boldsymbol{q}}_i]$. The same conditioning applied to Eq. (35) gives $\mathbb{E}[\boldsymbol{r}_i - \boldsymbol{b}_e^\star - \boldsymbol{\Gamma}_e^{\star\top} \tilde{\boldsymbol{q}}_i \mid e, \tilde{\boldsymbol{q}}_i, \boldsymbol{s}_i^\star] = \boldsymbol{s}_i^\star$, so the site-wise residual is unbiased at the subject level as well.

C Proof of Proposition 2

Proposition 2 (Margin Stability of the Site-aware Scaffold). *For edge $j \in \mathcal{I}$, let $r_{i,j}^{\mathrm{site}} = r_{i,j} - \hat{b}_{e,j} - \tilde{\boldsymbol{q}}_i^\top \hat{\boldsymbol{\gamma}}_{e,j}$ and $r_{i,j}^\star = r_{i,j} - b_{e,j}^\star - \tilde{\boldsymbol{q}}_i^\top \boldsymbol{\gamma}_{e,j}^\star$ denote the estimated and oracle site-wise residuals. Let $\mathcal{S}^\star$ be the scaffold obtained by applying the same selection rule, thresholds, and bootstrap draws to the oracle residuals, with oracle contrasts $d_{e,j}^\star$, $d_{\mathrm{com},j}^\star$, and $d_{\mathrm{com},j}^{\star(b)}$. Define the estimation perturbation and the oracle margin of edge j as*

$$\Delta_j = \max_{e\in\mathcal{E}_{\mathrm{tr}}} \sum_{y\in\{0,1\}} \sup_{i\in\mathcal{D}_e^{(y)}} \big|r_{i,j}^{\mathrm{site}} - r_{i,j}^\star\big|, \tag{38}$$

$$m_j^\star = \min\Big\{ \min_{e\in\mathcal{E}_{\mathrm{tr}}} |d_{e,j}^\star|,\ \min_{1\le b\le B} |d_{\mathrm{com},j}^{\star(b)}|,\ |d_{\mathrm{com},j}^\star|,\ \big||d_{\mathrm{com},j}^\star| - \tau_E\big| \Big\}. \tag{39}$$

If $\Delta_j < m_j^\star$, then $j \in \mathcal{S}$ if and only if $j \in \mathcal{S}^\star$.

Proof.

Fix an edge $j \in \mathcal{I}$, the source data, and the fitted and oracle deconfounder parameters. The empirical and oracle scaffolds share the standardized covariates, the thresholds $(\tau_E, \eta_E, \zeta_E)$, and the B site-bootstrap draws, so the argument is deterministic. We use the conventions of the main text: site-level contrasts are differences of Huber location estimates with a fixed threshold $\delta > 0$, consensus contrasts are component-wise medians, nonunique minimizers of either estimator are resolved by their midpoint, $\mathrm{sgn}(0) = 0$, and every source site contains both diagnostic groups.

We begin with the site-level contrasts. For a subject i at site e, the empirical and oracle residuals differ by

$$r_{i,j}^{\mathrm{site}} - r_{i,j}^\star = -\Big[(\hat{b}_{e,j} - b_{e,j}^\star) + \tilde{\boldsymbol{q}}_i^\top (\hat{\boldsymbol{\gamma}}_{e,j} - \boldsymbol{\gamma}_{e,j}^\star)\Big]. \tag{40}$$

For each class $y \in \{0,1\}$, define $a_{e,j}^{(y)} = \sup_{i \in \mathcal{D}_e^{(y)}} |r_{i,j}^{\mathrm{site}} - r_{i,j}^\star|$, so that $|r_{i,j}^{\mathrm{site}} - r_{i,j}^\star| \le a_{e,j}^{(y)}$ for every $i \in \mathcal{D}_e^{(y)}$, and write $\Delta_{e,j} = a_{e,j}^{(1)} + a_{e,j}^{(0)}$. By definition, $\Delta_j = \max_{e \in \mathcal{E}_{\mathrm{tr}}} \Delta_{e,j}$.

To pass from residuals to contrasts we need the Huber location estimator to be non-expansive under ℓ_∞ perturbations. For a nonempty finite sample $\boldsymbol{x} = (x_i)_{i \in A}$, let

$$\mathcal{M}_\delta(\boldsymbol{x}) = \arg\min_{\mu \in \mathbb{R}} \sum_{i \in A} \mathcal{H}_\delta(x_i - \mu), \qquad T_\delta(\boldsymbol{x}) = \tfrac{1}{2}\big(\min \mathcal{M}_\delta(\boldsymbol{x}) + \max \mathcal{M}_\delta(\boldsymbol{x})\big). \tag{41}$$

The objective is convex, continuous, and coercive, so $\mathcal{M}_\delta(\boldsymbol{x})$ is a nonempty compact interval. Since the Huber loss is differentiable, its minimizers are exactly the zeros of the score sum

$$\Psi_{\boldsymbol{x}}(\mu) = \sum_{i \in A} \psi_\delta(x_i - \mu), \qquad \psi_\delta(t) = \begin{cases} t, & |t| \le \delta, \\ \delta\,\mathrm{sgn}(t), & |t| > \delta, \end{cases} \tag{42}$$

which is continuous and non-increasing in μ with positive and negative limits as $\mu \to -\infty$ and $\mu \to +\infty$. Its zero set is therefore the interval with endpoints

$$\min \mathcal{M}_\delta(\boldsymbol{x}) = \inf\{\mu : \Psi_{\boldsymbol{x}}(\mu) \le 0\}, \qquad \max \mathcal{M}_\delta(\boldsymbol{x}) = \sup\{\mu : \Psi_{\boldsymbol{x}}(\mu) \ge 0\}. \tag{43}$$

Two properties of T_δ follow from this characterization. First, T_δ is monotone: if $x_i \le z_i$ for all $i \in A$, then ψ_δ being non-decreasing gives $\Psi_{\boldsymbol{x}}(\mu) \le \Psi_{\boldsymbol{z}}(\mu)$ for all μ, so both endpoints of the minimizer interval, and hence T_δ, can only increase,

$$T_\delta(\boldsymbol{x}) \le T_\delta(\boldsymbol{z}). \tag{44}$$

Second, T_δ is translation equivariant: shifting every observation by a constant c shifts the minimizer interval by c, so

$$T_\delta(\boldsymbol{x} + c\mathbf{1}) = T_\delta(\boldsymbol{x}) + c. \tag{45}$$

Now let $a = \|\boldsymbol{x} - \boldsymbol{x}^\star\|_\infty$, so that $\boldsymbol{x}^\star - a\mathbf{1} \le \boldsymbol{x} \le \boldsymbol{x}^\star + a\mathbf{1}$ coordinate-wise. Applying monotonicity to both inequalities and translation equivariance to the bounds gives $T_\delta(\boldsymbol{x}^\star) - a \le T_\delta(\boldsymbol{x}) \le T_\delta(\boldsymbol{x}^\star) + a$, that is,

$$|T_\delta(\boldsymbol{x}) - T_\delta(\boldsymbol{x}^\star)| \le \|\boldsymbol{x} - \boldsymbol{x}^\star\|_\infty. \tag{46}$$

Let $\mu_{e,j}^{(y)}$ and $\mu_{e,j}^{\star(y)}$ denote the Huber locations of the empirical and oracle residuals in class y at site e. Applying Eq. (46) to each class and using $d_{e,j} = \mu_{e,j}^{(1)} - \mu_{e,j}^{(0)}$ and $d_{e,j}^\star = \mu_{e,j}^{\star(1)} - \mu_{e,j}^{\star(0)}$,

$$|d_{e,j} - d_{e,j}^\star| \le |\mu_{e,j}^{(1)} - \mu_{e,j}^{\star(1)}| + |\mu_{e,j}^{(0)} - \mu_{e,j}^{\star(0)}| \le a_{e,j}^{(1)} + a_{e,j}^{(0)} = \Delta_{e,j} \le \Delta_j \tag{47}$$

for every $e \in \mathcal{E}_{\mathrm{tr}}$.

We next propagate this bound through the consensus and bootstrap contrasts, both of which are weighted medians of the site-level contrasts. For weights $w_e \ge 0$ with $\sum_e w_e > 0$, let

$$\mathcal{M}_w(\boldsymbol{x}) = \arg\min_{u \in \mathbb{R}} \sum_e w_e |x_e - u|, \qquad \mathrm{Med}_w(\boldsymbol{x}) = \tfrac{1}{2}\big(\min \mathcal{M}_w(\boldsymbol{x}) + \max \mathcal{M}_w(\boldsymbol{x})\big), \tag{48}$$

and let $F_{\boldsymbol{x},w}(u) = \sum_e w_e \mathbb{I}[x_e \le u] / \sum_e w_e$. The endpoints of the weighted-median interval are $\min \mathcal{M}_w(\boldsymbol{x}) = \inf\{u : F_{\boldsymbol{x},w}(u) \ge 1/2\}$ and $\max \mathcal{M}_w(\boldsymbol{x}) = \inf\{u : F_{\boldsymbol{x},w}(u) > 1/2\}$. If $x_e \le z_e$

Table 3: Demographic and clinical characteristics of the ABIDE, REST-meta-MDD, SRPBS, and ABCD datasets after site filtering. ASD, TD, MDD, ADHD, and HC denote autism spectrum disorder, typically developing, major depressive disorder, attention-deficit/hyperactivity disorder, and healthy control, respectively.

Characteristic	**ABIDE**	**REST-meta-MDD**	**SRPBS**	**ABCD (ADHD-task)**
Sample size	435	1,596	968	425
Class composition	209 ASD / 226 TD	845 MDD / 751 HC	484 MDD / 484 HC	213 ADHD / 212 HC
Age (mean $\pm$ std)	18.15 ± 9.53	39.28 ± 16.08	42.93 ± 13.58	9.47 ± 0.50
Sex (M/F)	367/68	629/967	437/531	246/179

for all e, then $F_{\boldsymbol{x},w}(u) \geq F_{\boldsymbol{z},w}(u)$ for every u, so both endpoints and their midpoint are coordinate-wise monotone, and translating the observations translates both endpoints. The sandwich argument used for Eq. (46) therefore gives

$$|\operatorname{Med}_w(\boldsymbol{x}) - \operatorname{Med}_w(\boldsymbol{x}^\star)| \leq \|\boldsymbol{x} - \boldsymbol{x}^\star\|_\infty. \tag{49}$$

The consensus contrast is the equal-weight median, $d_{\mathrm{com},j} = \operatorname{Med}_{\mathbf{1}}((d_{e,j})_e)$, so Eqs. (47) and (49) give

$$|d_{\mathrm{com},j} - d^\star_{\mathrm{com},j}| \leq \max_{e \in \mathcal{E}_{\mathrm{tr}}} |d_{e,j} - d^\star_{e,j}| \leq \Delta_j. \tag{50}$$

For bootstrap replicate b, let $w_e^{(b)}$ be the number of times site e is drawn, so that $\sum_e w_e^{(b)} = |\mathcal{E}_{\mathrm{tr}}|$; these multiplicities are the same in the empirical and oracle constructions and are unrelated to the deconfounder aggregation weights. The median of the resampled contrasts is the weighted median with these multiplicities, $d^{(b)}_{\mathrm{com},j} = \operatorname{Med}_{w^{(b)}}((d_{e,j})_e)$ and $d^{\star(b)}_{\mathrm{com},j} = \operatorname{Med}_{w^{(b)}}((d^\star_{e,j})_e)$, so the same argument gives

$$|d^{(b)}_{\mathrm{com},j} - d^{\star(b)}_{\mathrm{com},j}| \leq \Delta_j, \qquad b = 1, \ldots, B. \tag{51}$$

It remains to show that no quantity entering the selection rule crosses its threshold. Suppose $\Delta_j < m^\star_j$. Then Δ_j is strictly smaller than $|d^\star_{e,j}|$ for every e, than $|d^\star_{\mathrm{com},j}|$, and than $|d^{\star(b)}_{\mathrm{com},j}|$ for every b; in particular all of these oracle contrasts are nonzero. A real number within distance Δ_j of a nonzero x with $|x| > \Delta_j$ is nonzero and has the sign of x, so Eqs. (47), (50), and (51) yield

$$\operatorname{sgn}(d_{e,j}) = \operatorname{sgn}(d^\star_{e,j}), \quad \operatorname{sgn}(d_{\mathrm{com},j}) = \operatorname{sgn}(d^\star_{\mathrm{com},j}), \quad \operatorname{sgn}(d^{(b)}_{\mathrm{com},j}) = \operatorname{sgn}(d^{\star(b)}_{\mathrm{com},j}) \tag{52}$$

for all $e \in \mathcal{E}_{\mathrm{tr}}$ and $b = 1, \ldots, B$. Every indicator in the definitions of the sign consistency κ_j and the bootstrap stability π_j is therefore unchanged, and

$$\kappa_j = \kappa^\star_j, \qquad \pi_j = \pi^\star_j. \tag{53}$$

For the magnitude criterion, the reverse triangle inequality and Eq. (50) give

$$\big||d_{\mathrm{com},j}| - |d^\star_{\mathrm{com},j}|\big| \leq |d_{\mathrm{com},j} - d^\star_{\mathrm{com},j}| \leq \Delta_j < \big||d^\star_{\mathrm{com},j}| - \tau_E\big|. \tag{54}$$

The right-hand side is positive, so $|d^\star_{\mathrm{com},j}| \neq \tau_E$. If $|d^\star_{\mathrm{com},j}| > \tau_E$, then $|d_{\mathrm{com},j}| \geq |d^\star_{\mathrm{com},j}| - \Delta_j > \tau_E$; if $|d^\star_{\mathrm{com},j}| < \tau_E$, then $|d_{\mathrm{com},j}| \leq |d^\star_{\mathrm{com},j}| + \Delta_j < \tau_E$. Hence

$$\mathbb{I}\big[|d_{\mathrm{com},j}| > \tau_E\big] = \mathbb{I}\big[|d^\star_{\mathrm{com},j}| > \tau_E\big]. \tag{55}$$

Since the empirical and oracle scaffolds apply the same thresholds to the three criteria,

$$\begin{aligned} \mathbf{1}\{j \in \mathcal{S}\} &= \mathbb{I}\big[|d_{\mathrm{com},j}| > \tau_E,\ \kappa_j \geq \eta_E,\ \pi_j \geq \zeta_E\big] \\ &= \mathbb{I}\big[|d^\star_{\mathrm{com},j}| > \tau_E,\ \kappa^\star_j \geq \eta_E,\ \pi^\star_j \geq \zeta_E\big] = \mathbf{1}\{j \in \mathcal{S}^\star\}, \end{aligned} \tag{56}$$

which completes the proof.

D Datasets

D.1 Dataset Descriptions

Table 3 summarizes the demographic and clinical characteristics of the ABIDE, REST-meta-MDD, SRPBS, and ABCD datasets. Detailed per-environment statistics are provided in Tables 4–7.

Table 4: Statistics of the ABIDE dataset. T denotes the number of fMRI time points.

Site	N	TD	ASD	Age (mean $\pm$ std)	Male/Female	T
CALTECH	37	18	19	27.7 ± 10.3	29/8	146
KKI	48	28	20	10.0 ± 1.3	36/12	148
MAX_MUN	52	28	24	25.3 ± 11.8	48/4	140
OLIN	34	15	19	16.6 ± 3.4	29/5	206
PITT	56	27	29	18.9 ± 6.9	48/8	196
SBL	30	15	15	34.4 ± 8.5	30/0	196
SDSU	36	22	14	14.4 ± 1.8	29/7	176
STANFORD	39	20	19	10.0 ± 1.6	31/8	209
TRINITY	47	25	22	17.0 ± 3.4	47/0	146
YALE	56	28	28	12.7 ± 2.9	40/16	196
Total	**435**	**226**	**209**		**367/68**	

Table 5: Statistics of the REST-meta-MDD dataset. T denotes the number of fMRI time points.

Site	N	HC	MDD	Age (mean $\pm$ std)	Male/Female	T
S1	148	74	74	31.8 ± 8.5	63/85	200
S10	83	33	50	31.9 ± 11.0	42/41	202
S15	100	50	50	46.7 ± 16.4	42/58	230
S19	87	36	51	37.3 ± 11.0	32/55	230
S20	533	251	282	39.2 ± 14.7	186/347	232
S21	156	70	86	35.3 ± 12.6	69/87	230
S25	152	63	89	67.3 ± 6.7	50/102	230
S7	87	49	38	41.9 ± 12.6	34/53	174
S8	150	75	75	28.6 ± 10.9	57/93	190
S9	100	50	50	28.5 ± 8.6	54/46	190
Total	**1596**	**751**	**845**		**629/967**	

ABIDE. The Autism Brain Imaging Data Exchange (ABIDE) Di Martino et al. (2014) is a multi-site resting-state fMRI consortium for autism research. Starting from the quality-controlled cohort, we retain 10 sites that (i) include both individuals with autism spectrum disorder (ASD) and typically developing controls (TD), and (ii) provide sufficient samples per class for site-level evaluation. The resulting benchmark contains 435 subjects (209 ASD / 226 TD). The retained sites exhibit substantial heterogeneity in acquisition settings and cohort composition, including sequence length ($T = 140$–209 time points), site-level mean age (10.0–34.4 years), and sex distribution; SBL and TRINITY contain only male participants, whereas the other sites include both sexes. Table 4 summarizes the per-site statistics.

In addition to the standard AAL-116 setting, we evaluate ABIDE with the CC200 functional parcellation Craddock et al. (2012). The two settings share the same 435 subjects, diagnostic labels, scan sessions, and sites, and use identical training, validation, and test partitions in every fold, so the CC200 experiment isolates the effect of parcellation rather than introducing an additional cohort.

REST-meta-MDD. The REST-meta-MDD consortium Yan et al. (2019) aggregates resting-state fMRI data for major depressive disorder (MDD) across multiple clinical sites in China. We retain the 10 largest sites with adequate class representation, requiring each diagnostic class to contain at least 15 subjects per site. The resulting cohort contains 1,596 subjects (751 healthy controls [HC] / 845 MDD), including 629 males and 967 females. Site-level mean ages range from 28.5 to 67.3 years.

The retained cohort also exhibits substantial site-size imbalance. In particular, site S20 contains 533 subjects, approximately 33% of the retained cohort, whereas site S25 has a markedly older population with a mean age of 67.3 years. These characteristics provide substantial variation in population composition across evaluation environments. Table 5 summarizes the per-site statistics.

SRPBS. The Strategic Research Program for Brain Sciences (SRPBS) dataset Tanaka et al. (2021) is a multi-site, multi-disorder neuroimaging resource released through the DecNef consortium in Japan. We focus on the MDD cohort and retain 10 site/batch environments after quality control and cohort filtering, yielding 968 subjects (484 HC / 484 MDD), including 437 males and 531 females. Participants span a broad adult age range (18–80 years), with similar mean ages in the MDD and

Table 6: Statistics of the SRPBS dataset. T denotes the number of fMRI time points. For institutions with multiple acquisition batches, `batch1` and `batch2` denote distinct, non-overlapping subject cohorts, and each batch is treated as an independent environment.

Environment	N	HC	MDD	Age (mean $\pm$ std)	Male/Female	T
HUH-batch1	114	57	57	38.8 ± 13.1	56/58	143
HUH-batch2	146	73	73	40.2 ± 12.9	63/83	143
HRC	32	16	16	41.5 ± 12.0	10/22	143
HKH-batch1	56	28	28	45.9 ± 9.7	29/27	107
HKH-batch2	58	29	29	44.7 ± 9.9	29/29	107
COI-batch1	140	70	70	48.3 ± 13.4	57/83	240
COI-batch2	142	71	71	48.3 ± 12.7	57/85	240
KUT	32	16	16	40.9 ± 13.7	16/16	240
UTO-batch1	124	62	62	43.1 ± 14.4	54/70	240
UTO-batch2	124	62	62	36.3 ± 13.7	66/58	240
Total	**968**	**484**	**484**		**437/531**	

Table 7: Statistics of the ABCD dataset. T denotes the number of fMRI time points.

Site	N	HC	ADHD	Age (mean $\pm$ std)	Male/Female	T
site02	20	8	12	9.40 ± 0.49	13/7	383
site06	53	23	30	9.43 ± 0.50	30/23	383
site07	54	26	28	9.48 ± 0.50	34/20	383
site09	32	16	16	9.41 ± 0.49	20/12	383
site11	53	19	34	9.32 ± 0.47	29/24	383
site14	52	34	18	9.67 ± 0.46	22/30	383
site15	27	8	19	9.30 ± 0.46	16/11	383
site16	60	41	19	9.40 ± 0.49	33/27	383
site20	32	15	17	9.69 ± 0.46	17/15	383
site21	42	22	20	9.52 ± 0.50	32/10	383
Total	**425**	**212**	**213**		**246/179**	

HC groups (43.0 ± 13.5 and 43.0 ± 13.7 years, respectively), and the retained environments differ in acquisition length, with $T \in \{107, 143, 240\}$ time points.

Four institutions (HUH, HKH, COI, and UTO) contribute two data batches each, corresponding to the distinct acquisition protocol identifiers provided in the SRPBS release. The two batches of an institution contain non-overlapping subjects and are denoted `Institution-batch1` and `Institution-batch2` in Table 6. Because protocol differences within an institution induce shifts that are not captured by site identity alone, each site/batch unit is treated as an independent environment rather than being merged with the other batch from the same institution.

ABCD. The Adolescent Brain Cognitive Development (ABCD) Study® Casey et al. (2018) is a large-scale, multi-site longitudinal cohort designed to characterize brain development, cognition, and mental health from childhood into early adulthood. After quality control and site filtering, we retain 10 imaging sites with sufficient per-class coverage for attention-deficit/hyperactivity disorder (ADHD) versus HC classification. The resulting cohort contains 425 subjects (213 ADHD / 212 HC). All retained sessions contain $T = 383$ time points, and participants are narrowly concentrated around 9–10 years of age (mean 9.47 ± 0.50 years). Therefore, cross-site heterogeneity in this benchmark primarily reflects scanner/protocol variation, sex-ratio differences, and site-specific class composition, rather than broad age variation. Table 7 summarizes the per-site statistics.

Data used in the preparation of this article were obtained from the ABCD Study®, available through the NIMH Data Archive (NDA). The ABCD Study is supported by the National Institutes of Health and additional federal partners under multiple award numbers, including U01DA041048, U01DA050989, and U01DA051016, among others. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or

writing of this report. This manuscript reflects the authors' views and does not necessarily represent those of the NIH or ABCD consortium investigators.

D.2 Dataset Preprocessing

ABIDE. For ABIDE, we use resting-state fMRI and phenotypic data from the Autism Brain Imaging Data Exchange Di Martino et al. (2014), with preprocessed ROI time series obtained from the Preprocessed Connectomes Project (PCP). The standard setting uses the PCP-provided AAL-116 derivatives Tzourio-Mazoyer et al. (2002), and the ABIDE (CC200) setting uses the PCP-provided CC200 time series Craddock et al. (2012) produced by the same preprocessing pipeline. Subjects are retained only if they have valid diagnostic labels, site identifiers, the phenotypic variables required by the model, and usable ROI time series after quality control.

The two parcellation settings are matched exactly at the subject and scan levels: both contain the same 435 subjects, the same diagnostic labels and site identifiers, and the same scan sessions. Within every evaluation fold, the AAL-116 and CC200 settings also use identical training, validation, and test assignments. Thus, the atlas comparison controls the cohort and evaluation partitions while changing the ROI definition from 116 anatomical regions to 200 functionally defined regions. Subjects are retained only if they have valid diagnostic labels, site identifiers, phenotypic information required by the model, and usable ROI time series after quality control.

REST-meta-MDD. For REST-meta-MDD, we use consortium-provided derivatives that have undergone standardized local preprocessing rather than processing the raw imaging data ourselves Yan et al. (2019). We use AAL-116 ROI time series Tzourio-Mazoyer et al. (2002) together with diagnostic labels, site identifiers, and the demographic variables required by the model. Subjects without the required metadata or usable ROI time series are excluded.

SRPBS. For SRPBS, we use multi-site resting-state fMRI data and associated metadata released through the DecNef consortium Tanaka et al. (2021). Following the dataset-provided quality-control and preprocessing procedures, the rs-fMRI data are parcellated using the AAL-116 atlas Tzourio-Mazoyer et al. (2002). The mean BOLD signal within each ROI is extracted to form $\boldsymbol{X}_i \in \mathbb{R}^{T_i \times 116}$. The site/batch identifiers defined above are retained throughout the evaluation so that each of the 10 non-overlapping site/batch cohorts constitutes a distinct environment.

ABCD. For ABCD, we use data from the Adolescent Brain Cognitive Development Study® Casey et al. (2018), obtained through the NIMH Data Archive (Release 4.0). We focus on the baseline resting-state fMRI session (`baseline_year_1_arm_1`). Each subject's preprocessed rs-fMRI data are parcellated using the AAL-116 atlas Tzourio-Mazoyer et al. (2002), and the mean BOLD signal within each ROI is extracted to obtain $\boldsymbol{X}_i \in \mathbb{R}^{T_i \times 116}$. Subjects are retained only when the required diagnostic labels, site identifiers, demographic variables, and ROI time series are available.

E Baselines

E.1 Baseline Description

We compare CORE with different representative baselines across four categories: general GNNs, general OOD methods, graph OOD methods, brain-network models.

General GNNs.

- **GCN** Kipf & Welling (2016) is a spectral graph convolutional network derived from a first-order localized approximation of spectral graph convolutions, aggregating neighbor features via a symmetric normalized Laplacian.
- **GAT** Veličković et al. (2017) introduces masked self-attention layers so that each node learns adaptive weights over its neighbors, alleviating the uniform-aggregation limitation of GCN.
- **GIN** Xu et al. (2018) adopts a sum aggregator together with an injective update function, matching the expressive power of the Weisfeiler–Lehman graph isomorphism test.

General OOD methods.

- **IRM** Arjovsky et al. (2019) learns representations whose optimal classifier is invariant across training environments, penalizing environment-specific features for out-of-distribution generalization.
- **CORAL** Sun & Saenko (2016) is implemented here in a source-only form that aligns second-order feature statistics across the available source environments; no held-out-site statistics are used.

Graph OOD methods.

- **GSAT** Miao et al. (2022) is an interpretable and generalizable graph learning framework that, motivated by the information bottleneck principle, injects stochasticity into attention weights to select task-relevant subgraphs.
- **DisC** Fan et al. (2022) debiases GNNs under severely biased graph data by disentangling causal and bias substructures via an edge-mask generator and synthesizing counterfactual unbiased samples.
- **CEPG** Wang et al. (2026) promotes invariant information learning via environment-promoted training objectives, extracting features that remain predictive across diverse graph environments for OOD generalization.
- **DiSCO** Sun et al. (2026) employs a diverse and sparse mixture-of-experts over causal subgraphs, handling instance-level heterogeneous causal patterns for OOD graph learning.

Brain Networks.

- **STAGIN** Kim et al. (2021) learns dynamic connectome representations using spatio-temporal attention over sequences of functional graphs.
- **BrainNetTF** Kan et al. (2022) is a Transformer tailored to brain networks, using connection profiles as node features together with an orthonormal clustering readout aligned with ROI functional modules.
- **AGMGC** Noman et al. (2025) learns connectivity directly from raw fMRI time series through adaptive graph construction, and combines SplineCNN with multi-graph convolutions to capture localized and global dependencies for brain disorder classification.
- **FC-HGNN** Gu et al. (2025) is a two-stage end-to-end heterogeneous graph neural network built on brain functional connectivity, applying type-aware message passing over a heterogeneous brain graph to identify mental disorders.
- **XG-GNN** Qiu et al. (2024) combines meta-learning with explainability-generalizable regularization to learn domain-generalizable diagnostic GNNs and explanation factors from fMRI BOLD signals.
- **LHDFormer** Xue et al. (2025) models long-range and high-order dependencies in time-varying fMRI representations using adaptive embeddings and Transformer-based temporal modeling.
- **NeuroH-TGL** Li et al. (2025b) learns temporal brain graphs while accounting for neuro-heterogeneity across subjects.
- **BrainOOD** Xu et al. (2025) is an OOD-generalizable brain network framework that couples a feature selector, a structure extractor, and an improved graph information bottleneck objective to recover causal subgraphs predictive across unseen sites.
- **DeCI** Yu et al. (2026) moves beyond static functional connectivity by directly modeling raw fMRI time series, using cycle-and-drift decomposition with channel-independence for brain disorder classification.

F Implementation Details

We implement CORE in PyTorch and conduct all experiments on NVIDIA GeForce RTX 5090 GPUs. The observed covariates $\boldsymbol{q}_i$ comprise demographic variables (age and sex), head motion (mean framewise displacement), and acquisition variables (sequence length and TR), i.e., $d = 5$; missing

entries are imputed with the pooled training median. All covariates are standardized with pooled training statistics, and site-specific coefficients are fitted by Huber regression with $\delta = 1.345$ using IRLS. Bootstrap sign stability is estimated with $B = 50$ site resamples. Within each site the site-level covariate columns coincide with the intercept column; the IRLS updates use the pseudoinverse, and the site-wise residuals do not depend on how the intercept is apportioned among these columns. We use a two-layer GCN Kipf & Welling (2016) with hidden dimension 64 as the backbone, with gate prior strength $\lambda = 0.3$ and gate temperature $\tau = 1$, optimized using Adam with a learning rate of 1×10^{-3}, weight decay of 5×10^{-4}, batch size 64, and 100 epochs. Sliding-window FC is computed with window length $W = 30$ and stride $S_w = 5$ time points, so window duration scales with each site's TR. The target gate mass is set to $K = 80$ for ABIDE, REST-meta-MDD, and SRPBS, and $K = 150$ for ABCD, and the gate-budget regularization weight is $\beta = 5 \times 10^{-4}$. In each LOSO fold, one site is held out for testing, two randomly selected source sites are reserved for validation, and the remaining sites are used for training. Site-specific deconfounders, scaffold statistics, source-reliability scores, and aggregation weights are estimated using only the training sites; validation sites are excluded from $\mathcal{E}_{\text{tr}}$ and are therefore residualized with the frozen aggregate deconfounder, matching the test-time condition. Hyperparameters are tuned and the training epoch is selected on the validation sites. The scaffold thresholds are fixed to τ_E at the 80th percentile of the absolute consensus contrasts $\{|d_{\text{com},j}|\}_{j\in\mathcal{I}}$ computed from the training sites, $\eta_E = 0.75$, and $\zeta_E = 0.70$ in all experiments; these values were chosen by grid search on the validation sites over the 70th–90th percentiles and $\eta_E, \zeta_E \in \{0.70, 0.75, 0.80, 0.85, 0.90\}$. Results are reported as mean $\pm$ standard deviation across held-out test sites.

G Complexity Analysis

Here we analyze the computational complexity of the proposed CORE. The computational cost is mainly dominated by static FC computation, scaffold-based transient profiling, and gated message passing on the line graph. Let N, P, and T denote the number of subjects, ROIs, and time points, respectively. Let $M = \mathcal{O}(P^2)$ be the number of full FC edges, $M_S = |S|$ be the number of scaffold edges, and $R = \lfloor (T - W)/S_w \rfloor + 1$ be the number of sliding windows. B_{sz} is the mini-batch size, D is the embedding dimension, L is the number of GNN layers, and N_m is the maximum number of line-graph neighbors per scaffold edge. Static FC computation and scaffold extraction take $\mathcal{O}(NTP^2 + NM)$ time complexity, while transient profiling over scaffold edges introduces an additional cost of $\mathcal{O}(NRWM_S)$. During training, gated message passing on the sparse line graph takes $\mathcal{O}(LB_{\text{sz}}M_S N_m D)$ per mini-batch. As a result, for N_{ep} training epochs, the overall computational complexity of CORE is $\mathcal{O}\left(NTP^2 + NM + NRWM_S + N_{\text{ep}}LNM_S N_m D\right)$. Since N_m is small in practice, the trainable cost scales approximately linearly with the scaffold size M_S.

H Algorithm

The overall training and inference process of the proposed CORE is shown in Algorithm 1.

I More Experimental Results

I.1 ID Performance Comparison

Although CORE is designed for cross-site OOD generalization, we also evaluate its predictive performance under in-distribution (ID) settings. Within each dataset, subjects from all sites are pooled and partitioned into 10 folds stratified by diagnostic label; in each of the 10 runs, one fold is held out for testing, the next two folds are used for validation, and the remaining seven folds are used for training, yielding a 7:2:1 split without holding out entire imaging sites as in the LOSO evaluation. Since every site is represented in the training folds, site-specific deconfounders and scaffold statistics are estimated for all sites from the training subjects of each run, and validation and test subjects are residualized with the frozen aggregate deconfounder as at inference in the LOSO evaluation. As shown in Table 8, CORE achieves the highest mean AUC across all five evaluated settings and the highest mean ACC on REST-meta-MDD, SRPBS, and ABCD. It ranks second in ACC under both ABIDE parcellations, where DeCI achieves the highest mean. While the relative performance of the baselines varies across datasets and metrics, CORE remains consistently competitive across

Algorithm 1 Training and inference of CORE

Require: Source datasets $\{\widehat{\mathcal{D}}^e\}_{e\in\mathcal{E}_{\mathrm{tr}}}$, test inputs $\{(\boldsymbol{X}_i, \boldsymbol{q}_i)\}_{i=1}^{N_{\mathrm{test}}}$, temporal parameters (W, S_w), scaffold parameters $(\tau_E, \eta_E, \zeta_E, B)$, gate parameters (λ, τ, K), loss weight β, and GNN depth L.
Ensure: Test predictions $\widehat{\boldsymbol{Y}}_{\mathrm{test}}$.

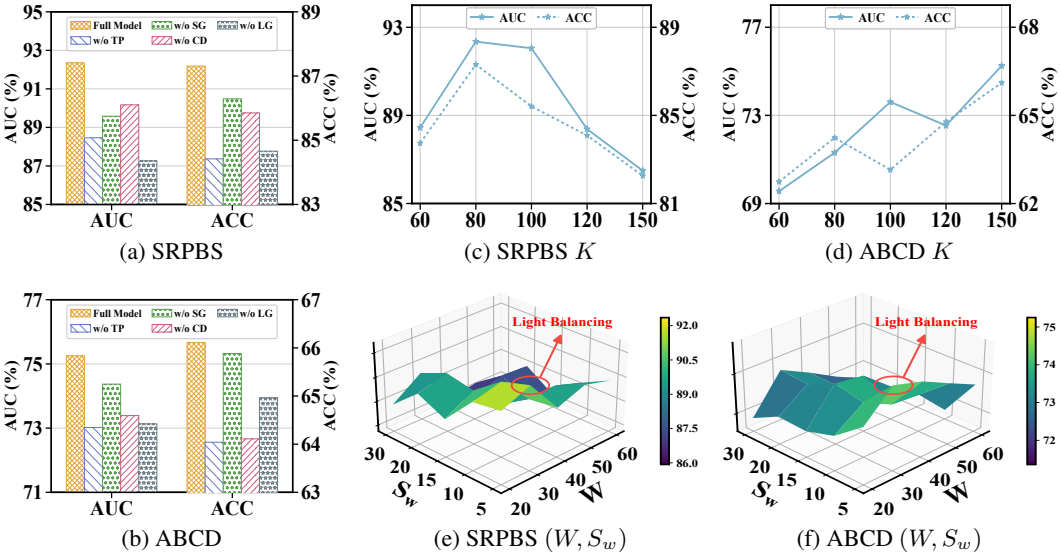
Define $\widehat{\mathcal{D}}_{\mathrm{tr}} = \bigcup_{e\in\mathcal{E}_{\mathrm{tr}}} \widehat{\mathcal{D}}^e$.
Stage 1: Site-Aware Confounder Decoupling and Scaffold Extraction
Estimate covariate means $\boldsymbol{\mu}_q$ and standard deviations $\boldsymbol{\sigma}_q$ from $\widehat{\mathcal{D}}_{\mathrm{tr}}$.
for each $e \in \mathcal{E}_{\mathrm{tr}}$ **do**
  Compute full-scan static FC vectors $\boldsymbol{r}_i$ and standardized covariates $\tilde{\boldsymbol{q}}_i = (\boldsymbol{q}_i - \boldsymbol{\mu}_q)/\boldsymbol{\sigma}_q$ for subjects at site e.
  Estimate $(\hat{\boldsymbol{b}}_e, \hat{\boldsymbol{\Gamma}}_e, \hat{\boldsymbol{\theta}}_e)$ by Huber regression (Eq. (5)); $\hat{\boldsymbol{\theta}}_e$ is not subtracted.
  Obtain $\boldsymbol{r}_i^{\mathrm{site}}$ (Eq. (6)).
  Compute the site-level contrast $\boldsymbol{d}_e$ from class-wise robust means of $\boldsymbol{r}_i^{\mathrm{site}}$.
end for
Compute $\boldsymbol{d}_{\mathrm{com}}$ as the component-wise median of $\{\boldsymbol{d}_e\}_{e\in\mathcal{E}_{\mathrm{tr}}}$.
Compute κ_j and estimate π_j using B source-site bootstrap replicates (Eq. (10)).
Extract the population scaffold $\mathcal{S}$ (Eq. (11)).
Compute disagreement scores u_e and reliability weights α_e (Eq. (7)).
Form the weighted deconfounder $(\bar{\boldsymbol{b}}_w, \bar{\boldsymbol{\Gamma}}_w)$ (Eq. (8)).
Stage 2: Transient Pathway Profiling
Fix an enumeration $(j_p)_{p=1}^{M_{\mathcal{S}}}$ of $\mathcal{S}$, where $M_{\mathcal{S}} = |\mathcal{S}|$.
for each training subject i **do**
  Compute windowed correlations $\rho_{i,j}^{(t)}$ and Fisher-z values $c_{i,j}^{(t)}$ for $j \in \mathcal{S}$ using (W, S_w).
  Compute temporal descriptors $s_{i,j}$ and $f_{i,j}$ (Eq. (14)).
  Construct $\boldsymbol{h}_{i,p}^{(0)}$ from $\boldsymbol{r}_i^{\mathrm{site}}$ and the temporal descriptors (Eq. (15)).
end for
Compute prior scores $\boldsymbol{\mu}_0$, line-graph adjacency $\boldsymbol{A}$, and propagation matrix $\tilde{\boldsymbol{A}}$.
Stage 3: Prior-Guided Subject-Adaptive Gating and Training
Initialize trainable GNN and MLP parameters; keep source-derived statistics, $\mathcal{S}$, $\boldsymbol{\mu}_0$, and $\tilde{\boldsymbol{A}}$ fixed.

while not converged **do**
  Sample a batch $\mathcal{B} \subset \widehat{\mathcal{D}}_{\mathrm{tr}}$.
  for each subject i in $\mathcal{B}$ **do**
    Compute $\boldsymbol{c}_i$, $g_{i,p}$, and $\boldsymbol{h}_{i,p}^{(L)}$ (Eqs. (18)–(20)).
    Obtain $\boldsymbol{z}_i$ by gated readout (Eq. (21)) and predict $\hat{y}_i = \rho(\boldsymbol{z}_i)$.
  end for
  Update trainable parameters using the batch loss $\mathcal{L} = \mathcal{L}_{\mathrm{cls}} + \beta\mathcal{L}_{\mathrm{budget}}$ (Eq. (24)).
end while
Select the parameters with the best validation score, where validation subjects are processed as in Stage 4.
Stage 4: Inference on Held-out Sites
for each test subject $i = 1, \ldots, N_{\mathrm{test}}$ **do**
  Compute $\boldsymbol{r}_i$ from $\boldsymbol{X}_i$ and standardize $\boldsymbol{q}_i$ using the source statistics $(\boldsymbol{\mu}_q, \boldsymbol{\sigma}_q)$.
  Obtain $\boldsymbol{r}_i^{\mathrm{res}}$ using the frozen deconfounder $(\bar{\boldsymbol{b}}_w, \bar{\boldsymbol{\Gamma}}_w)$ (Eq. (8)).
  Compute temporal descriptors on $\mathcal{S}$ using (W, S_w) and construct $\boldsymbol{h}_{i,p}^{(0)}$ with $\boldsymbol{r}_i^{\mathrm{res}}$ (Eq. (15)).
  Compute $\boldsymbol{c}_i$, $g_{i,p}$, and $\boldsymbol{h}_{i,p}^{(L)}$ using the trained model and fixed $(\boldsymbol{\mu}_0, \tilde{\boldsymbol{A}})$.
  Obtain $\boldsymbol{z}_i$ by gated readout and predict $\hat{y}_i = \rho(\boldsymbol{z}_i)$.
end for
return $\widehat{\boldsymbol{Y}}_{\mathrm{test}} = (\hat{y}_i)_{i=1}^{N_{\mathrm{test}}}$.

the evaluated ASD, MDD, and ADHD-related tasks. These results demonstrate that the proposed framework is also effective for conventional subject-level prediction. Together with the LOSO results

Table 8: Results on ABIDE, ABIDE (CC200), REST-meta-MDD, SRPBS, and ABCD (ADHD-task) under in-distribution (ID) settings. **Bold** results indicate the best performance, underlined indicates second best.

Type	Method	ABIDE AUC	ABIDE ACC	ABIDE (CC200) AUC	ABIDE (CC200) ACC	REST-meta-MDD AUC	REST-meta-MDD ACC	SRPBS AUC	SRPBS ACC	ABCD (ADHD-task) AUC	ABCD (ADHD-task) ACC
GNN	GCN	58.90$_{\pm 9.10}$	55.80$_{\pm 10.00}$	48.34$_{\pm 9.12}$	51.05$_{\pm 7.95}$	61.91$_{\pm 3.44}$	56.41$_{\pm 2.19}$	83.23$_{\pm 4.21}$	74.87$_{\pm 5.19}$	70.13$_{\pm 7.51}$	64.45$_{\pm 7.47}$
	GAT	60.10$_{\pm 6.80}$	58.90$_{\pm 3.70}$	59.12$_{\pm 9.54}$	55.88$_{\pm 8.41}$	62.91$_{\pm 3.44}$	58.60$_{\pm 2.89}$	81.02$_{\pm 4.87}$	69.58$_{\pm 5.42}$	70.65$_{\pm 5.77}$	62.31$_{\pm 6.25}$
	GIN	58.50$_{\pm 5.40}$	57.50$_{\pm 5.60}$	53.76$_{\pm 8.21}$	48.91$_{\pm 11.23}$	58.11$_{\pm 2.97}$	55.41$_{\pm 1.82}$	81.62$_{\pm 4.77}$	70.39$_{\pm 5.61}$	72.27$_{\pm 6.01}$	63.25$_{\pm 5.97}$
OOD	CORAL	55.90$_{\pm 6.60}$	55.10$_{\pm 6.10}$	55.67$_{\pm 9.38}$	50.21$_{\pm 8.76}$	58.85$_{\pm 3.52}$	56.07$_{\pm 3.46}$	79.82$_{\pm 4.71}$	69.33$_{\pm 3.86}$	70.96$_{\pm 6.80}$	61.66$_{\pm 7.12}$
	IRM	55.40$_{\pm 7.80}$	54.20$_{\pm 6.10}$	50.12$_{\pm 11.45}$	54.89$_{\pm 6.32}$	59.62$_{\pm 2.57}$	56.26$_{\pm 3.08}$	79.14$_{\pm 4.43}$	69.19$_{\pm 4.52}$	70.73$_{\pm 6.70}$	63.98$_{\pm 8.24}$
Graph OOD	GSAT	56.20$_{\pm 4.90}$	53.80$_{\pm 6.30}$	49.88$_{\pm 7.82}$	51.34$_{\pm 6.05}$	57.65$_{\pm 3.84}$	54.35$_{\pm 3.97}$	81.69$_{\pm 3.25}$	70.87$_{\pm 4.62}$	70.92$_{\pm 7.76}$	63.78$_{\pm 6.62}$
	DisC	58.80$_{\pm 7.20}$	54.40$_{\pm 7.30}$	54.19$_{\pm 6.22}$	52.11$_{\pm 6.87}$	58.75$_{\pm 4.21}$	56.44$_{\pm 3.17}$	81.91$_{\pm 4.39}$	72.46$_{\pm 3.83}$	72.27$_{\pm 6.08}$	63.54$_{\pm 6.60}$
	CEPG	50.19$_{\pm 6.63}$	50.57$_{\pm 7.38}$	44.52$_{\pm 10.15}$	49.33$_{\pm 9.02}$	56.36$_{\pm 4.72}$	54.26$_{\pm 4.15}$	74.67$_{\pm 8.22}$	67.74$_{\pm 5.83}$	71.24$_{\pm 8.20}$	63.74$_{\pm 9.19}$
	DiSCO	52.27$_{\pm 6.19}$	51.72$_{\pm 5.39}$	51.98$_{\pm 8.65}$	50.04$_{\pm 7.21}$	59.41$_{\pm 4.86}$	55.77$_{\pm 2.87}$	84.14$_{\pm 3.26}$	72.78$_{\pm 4.39}$	68.75$_{\pm 9.15}$	63.14$_{\pm 8.33}$
Brain Networks	STAGIN	59.42$_{\pm 6.15}$	57.38$_{\pm 5.24}$	58.15$_{\pm 7.22}$	56.41$_{\pm 6.89}$	59.18$_{\pm 4.25}$	55.72$_{\pm 3.41}$	81.33$_{\pm 5.12}$	71.45$_{\pm 4.98}$	69.21$_{\pm 6.84}$	60.15$_{\pm 6.11}$
	BrainNetTF	60.60$_{\pm 7.00}$	58.90$_{\pm 3.70}$	58.91$_{\pm 10.22}$	51.45$_{\pm 8.91}$	61.70$_{\pm 2.40}$	56.92$_{\pm 2.31}$	78.43$_{\pm 7.78}$	69.39$_{\pm 6.61}$	70.08$_{\pm 7.30}$	58.34$_{\pm 7.89}$
	XG-GNN	52.61$_{\pm 7.51}$	53.08$_{\pm 3.04}$	49.25$_{\pm 13.02}$	55.31$_{\pm 8.44}$	59.31$_{\pm 3.99}$	55.33$_{\pm 3.17}$	83.43$_{\pm 7.34}$	72.21$_{\pm 8.26}$	71.32$_{\pm 8.35}$	56.90$_{\pm 5.94}$
	AGMGC	58.04$_{\pm 7.33}$	52.41$_{\pm 6.02}$	54.83$_{\pm 7.94}$	52.88$_{\pm 5.56}$	57.78$_{\pm 5.51}$	53.88$_{\pm 1.75}$	85.64$_{\pm 3.91}$	75.83$_{\pm 3.15}$	69.75$_{\pm 6.99}$	61.65$_{\pm 7.85}$
	FC-HGNN	55.43$_{\pm 11.85}$	51.74$_{\pm 5.76}$	56.77$_{\pm 9.81}$	54.12$_{\pm 5.99}$	60.27$_{\pm 5.67}$	55.95$_{\pm 4.12}$	76.07$_{\pm 5.48}$	67.22$_{\pm 6.21}$	72.69$_{\pm 7.92}$	61.68$_{\pm 9.03}$
	LHDFormer	60.85$_{\pm 5.41}$	57.92$_{\pm 4.77}$	61.34$_{\pm 5.18}$	58.05$_{\pm 5.02}$	62.15$_{\pm 3.86}$	57.84$_{\pm 2.95}$	84.72$_{\pm 4.35}$	74.55$_{\pm 3.82}$	71.84$_{\pm 6.05}$	62.91$_{\pm 5.44}$
	NeuroH-TGL	59.76$_{\pm 7.12}$	58.11$_{\pm 5.88}$	60.52$_{\pm 6.45}$	57.29$_{\pm 5.61}$	61.42$_{\pm 3.92}$	58.21$_{\pm 3.04}$	83.91$_{\pm 4.67}$	73.18$_{\pm 4.15}$	70.93$_{\pm 6.32}$	61.85$_{\pm 6.72}$
	BrainOOD	61.29$_{\pm 8.10}$	58.19$_{\pm 6.64}$	62.45$_{\pm 4.95}$	59.08$_{\pm 4.25}$	63.49$_{\pm 4.01}$	60.01$_{\pm 3.11}$	85.11$_{\pm 4.02}$	75.92$_{\pm 4.14}$	72.17$_{\pm 6.46}$	64.92$_{\pm 5.83}$
	DeCI	59.89$_{\pm 4.63}$	**62.07**$_{\pm 3.65}$	57.94$_{\pm 4.52}$	**62.27**$_{\pm 3.81}$	53.85$_{\pm 2.94}$	54.39$_{\pm 2.50}$	86.63$_{\pm 5.71}$	78.29$_{\pm 4.68}$	60.00$_{\pm 7.86}$	60.72$_{\pm 4.03}$
	CORE	**63.35**$_{\pm 7.82}$	60.12$_{\pm 6.05}$	**64.91**$_{\pm 7.42}$	61.52$_{\pm 4.51}$	**64.16**$_{\pm 3.12}$	**60.41**$_{\pm 2.85}$	**87.23**$_{\pm 3.01}$	**79.58**$_{\pm 2.45}$	**73.05**$_{\pm 6.12}$	**65.94**$_{\pm 6.33}$



(a) SRPBS (c) SRPBS K (d) ABCD K

(b) ABCD (e) SRPBS (W, S_w) (f) ABCD (W, S_w)

Figure 6: Ablation and sensitivity analyses on SRPBS and ABCD: (a,b) component ablations; (c,d) sensitivity to gate budget K; (e,f) sensitivity to window length W and stride S_w.

in the main text, the ID evaluation shows that CORE combines improved unseen-site generalization with strong in-distribution classification performance.

I.2 More Ablation Study

We further present ablation results of CORE under OOD settings on SRPBS and ABCD in Fig. 6(a,b). On SRPBS, the full model achieves the highest mean AUC and ACC. Removing TP reduces performance, indicating that scaffold-level temporal descriptors provide complementary predictive information beyond static FC. Replacing site-aware deconfounding with global deconfounding also lowers performance, supporting the importance of accounting for site-dependent confounder effects. The declines observed for w/o LG and w/o SG highlight the contributions of pathway interactions and subject-adaptive weighting, respectively. Specifically, line-graph message passing models interactions among scaffold pathways sharing an ROI, while gating adjusts their contributions to individual predictions. On ABCD, the full model again achieves the highest mean AUC, with all ablated variants showing lower performance. The decreases for w/o TP and w/o CD support the contributions of temporal information and site-aware deconfounding, while w/o LG further demonstrates the benefit

Table 9: Controlled comparison of temporal representation and segmentation choices in CORE on ABIDE under AAL and CC200 parcellations.

Temporal variant	ABIDE		ABIDE (CC200)	
	AUC	ACC	AUC	ACC
CORE-STD	$64.63_{\pm 7.02}$	$62.28_{\pm 6.12}$	$62.85_{\pm 8.44}$	$61.82_{\pm 8.94}$
CORE-MASD	$64.78_{\pm 6.97}$	$62.41_{\pm 6.08}$	$62.94_{\pm 8.38}$	$61.95_{\pm 8.91}$
CORE-GRU	$\mathbf{65.71}_{\pm 6.85}$	$62.77_{\pm 5.99}$	$63.49_{\pm 8.29}$	$61.96_{\pm 8.87}$
CORE-CP	$65.22_{\pm 6.91}$	$63.05_{\pm 5.98}$	$63.31_{\pm 8.33}$	$62.15_{\pm 8.89}$
CORE	$65.65_{\pm 6.42}$	$\mathbf{64.83}_{\pm 6.62}$	$\mathbf{64.61}_{\pm 7.12}$	$\mathbf{62.18}_{\pm 8.35}$

Table 10: Controlled ablation of the scaffold prior and subject-conditioned term in the gate logits on ABIDE under AAL and CC200 parcellations.

Gating variant	ABIDE		ABIDE (CC200)	
	AUC	ACC	AUC	ACC
Prior-only	$62.67_{\pm 6.34}$	$60.83_{\pm 5.51}$	$61.14_{\pm 7.88}$	$59.86_{\pm 8.42}$
Subject-only	$64.12_{\pm 7.18}$	$62.19_{\pm 6.12}$	$62.51_{\pm 8.54}$	$61.42_{\pm 9.06}$
CORE	$\mathbf{65.65}_{\pm 6.42}$	$\mathbf{64.83}_{\pm 6.62}$	$\mathbf{64.61}_{\pm 7.12}$	$\mathbf{62.18}_{\pm 8.35}$

of modeling pathway interactions rather than encoding scaffold nodes independently. For ACC, w/o TP and w/o CD also fall below the full model, whereas w/o SG remains close to it. Thus, the benefit of subject-adaptive gating is more evident in AUC than in ACC on ABCD, while the contributions of individual components vary across datasets and metrics.

Additionally, we evaluate four variants of transient pathway profiling on ABIDE under the AAL and CC200 parcellations. CORE-STD retains temporal volatility but removes temporal range, whereas CORE-MASD replaces range with the mean absolute successive difference between consecutive windowed FC estimates, incorporating local temporal ordering while preserving feature dimensionality. CORE-GRU replaces both handcrafted descriptors with a GRU encoding of the FC trajectory on each scaffold edge, while CORE-CP replaces fixed sliding windows with change-point-defined segments. As shown in Table 9, combining volatility and range consistently outperforms volatility alone across both metrics and parcellations, supporting the complementary contribution of temporal range. MASD provides only marginal improvements over volatility alone and remains below the volatility–range configuration. The GRU encoder achieves a slightly higher AUC under AAL but lower ACC, and trails the reference configuration on both metrics under CC200. Change-point segmentation also yields lower mean performance under both parcellations. Thus, incorporating local temporal ordering, sequence encoding, or adaptive segmentation does not provide consistent gains over the volatility–range configuration. These results support the selected descriptors as a compact temporal representation for the evaluated cross-site prediction tasks.

We next examine the contributions of population guidance and subject-conditioned adaptation to the gating mechanism on ABIDE under the AAL and CC200 parcellations. Prior-only retains the scaffold-derived prior but removes the subject-conditioned term from the gate logits, whereas Subject-only removes the prior term and computes gates from local pathway features and global context. As shown in Table 10, Subject-only outperforms Prior-only under both parcellations, highlighting the value of adapting pathway weights to individual connectivity features. Combining the two terms further improves performance, yielding the highest mean AUC and ACC under both parcellations. These results support the complementary roles of scaffold-derived population guidance and subject-conditioned pathway weighting in cross-site prediction.

Finally, we compare CORE w/o TP with representative static-FC baselines on four datasets under LOSO evaluation. This variant removes temporal descriptors. As shown in Table 11, CORE w/o TP achieves the highest mean AUC and ACC on ABIDE and REST-meta-MDD. On SRPBS, it attains the highest mean ACC but trails BrainOOD in AUC, whereas on ABCD, it leads in AUC and ranks second in ACC. These results show that the retained framework remains competitive across datasets using static FC, indicating that its predictive performance is not solely attributable to additional temporal information.

Table 11: Comparison of CORE w/o TP with representative static-FC baselines under LOSO evaluation. **Bold** denotes the best performance.

Method	ABIDE		REST-meta-MDD		SRPBS		ABCD	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
DisC	$56.67_{\pm 7.85}$	$56.14_{\pm 8.68}$	$56.21_{\pm 4.42}$	$53.28_{\pm 3.88}$	$73.72_{\pm 19.78}$	$69.97_{\pm 16.93}$	$69.59_{\pm 7.70}$	$58.83_{\pm 9.45}$
BrainNetTF	$62.85_{\pm 6.20}$	$60.90_{\pm 7.06}$	$61.50_{\pm 7.40}$	$58.20_{\pm 5.60}$	$61.56_{\pm 17.03}$	$60.41_{\pm 10.80}$	$68.28_{\pm 4.54}$	$56.91_{\pm 10.43}$
FC-HGNN	$55.26_{\pm 7.90}$	$52.23_{\pm 4.58}$	$58.69_{\pm 6.69}$	$56.78_{\pm 5.71}$	$57.30_{\pm 10.60}$	$53.80_{\pm 6.40}$	$71.08_{\pm 5.74}$	$59.10_{\pm 12.34}$
BrainOOD	$63.82_{\pm 6.73}$	$59.66_{\pm 9.70}$	$63.22_{\pm 6.29}$	$58.45_{\pm 5.20}$	$\mathbf{90.00}_{\pm 7.33}$	$77.70_{\pm 15.23}$	$72.77_{\pm 8.07}$	$\mathbf{64.36}_{\pm 7.24}$
CORE w/o TP	$\mathbf{65.12}_{\pm 7.91}$	$\mathbf{62.08}_{\pm 8.84}$	$\mathbf{65.95}_{\pm 9.04}$	$\mathbf{64.12}_{\pm 5.56}$	$88.46_{\pm 9.61}$	$\mathbf{84.42}_{\pm 11.24}$	$\mathbf{73.02}_{\pm 9.02}$	$64.04_{\pm 8.65}$

Table 12: Sensitivity to scaffold-selection criteria on ABIDE under AAL and CC200 parcellations. **Bold** settings denote defaults.

Criterion	Setting	ABIDE (AAL)		ABIDE (CC200)	
		AUC	ACC	AUC	ACC
Percentile for τ_E	70th	$63.77_{\pm 7.12}$	$61.29_{\pm 6.24}$	$62.11_{\pm 8.51}$	$60.85_{\pm 9.02}$
	75th	$64.82_{\pm 6.95}$	$62.65_{\pm 6.08}$	$63.15_{\pm 8.33}$	$61.92_{\pm 8.91}$
	80th	$65.65_{\pm 6.42}$	$64.83_{\pm 6.62}$	$64.61_{\pm 7.12}$	$62.18_{\pm 8.35}$
	85th	$65.21_{\pm 6.91}$	$63.12_{\pm 5.96}$	$63.55_{\pm 8.29}$	$62.45_{\pm 8.85}$
	90th	$64.55_{\pm 7.05}$	$62.88_{\pm 6.11}$	$63.02_{\pm 8.42}$	$61.88_{\pm 8.98}$
Sign consistency η_E	0.70	$65.15_{\pm 6.92}$	$63.22_{\pm 5.94}$	$63.45_{\pm 8.31}$	$62.35_{\pm 8.88}$
	0.75	$65.65_{\pm 6.42}$	$64.83_{\pm 6.62}$	$64.61_{\pm 7.12}$	$62.18_{\pm 8.35}$
	0.90	$64.41_{\pm 7.11}$	$62.15_{\pm 6.14}$	$62.85_{\pm 8.45}$	$61.70_{\pm 9.06}$
Bootstrap stability ζ_E	**0.70**	$65.65_{\pm 6.42}$	$64.83_{\pm 6.62}$	$64.61_{\pm 7.12}$	$62.18_{\pm 8.35}$
	0.75	$65.41_{\pm 6.90}$	$63.32_{\pm 5.93}$	$63.58_{\pm 8.28}$	$62.55_{\pm 8.85}$
	0.80	$65.25_{\pm 6.93}$	$63.10_{\pm 5.97}$	$63.42_{\pm 8.32}$	$62.20_{\pm 8.89}$
	0.85	$64.88_{\pm 7.01}$	$62.75_{\pm 6.06}$	$63.10_{\pm 8.38}$	$61.85_{\pm 8.95}$
	0.90	$64.35_{\pm 7.09}$	$62.25_{\pm 6.17}$	$62.65_{\pm 8.49}$	$61.35_{\pm 9.01}$

I.3 More Sensitivity Study

We also present the sensitivity study of CORE under OOD settings on the SRPBS and ABCD datasets. As shown in Fig. 6(c), SRPBS exhibits a clear inverted-U trend as K increases. Performance peaks around $K = 80$, while both smaller and larger budgets degrade results. This pattern suggests that a small budget may overly suppress informative pathways, whereas a large budget may increase the contributions of less informative pathways. On ABCD, Fig. 6(d) shows that performance generally improves with increasing K, with the best result at the largest evaluated budget. This suggests that ABCD benefits from broader pathway coverage, likely due to greater inter-subject variability and more distributed diagnostic patterns. Nevertheless, small budgets still underperform, indicating that insufficient pathway coverage limits subject-adaptive message passing. For temporal profiling, Fig. 6(e,f) shows that both SRPBS and ABCD favor moderate windowing with relatively fine temporal strides. Coarse temporal partitioning degrades performance, indicating that transient connectivity dynamics are informative for cross-site generalization, while overly aggressive segmentation may yield unstable dynamic FC estimates.

We further examine sensitivity to the scaffold-selection criteria on ABIDE under AAL and CC200 parcellations. The contrast threshold τ_E, sign-consistency cutoff η_E, and bootstrap-stability cutoff ζ_E are varied independently, with the other two fixed at their default values determined through source-side validation. The default configuration uses the 80th percentile of the absolute consensus contrasts, $\eta_E = 0.75$, and $\zeta_E = 0.70$. We vary the contrast percentile from the 70th to the 90th in increments of five and evaluate $\zeta_E \in \{0.70, 0.75, 0.80, 0.85, 0.90\}$. Since each fold has seven training sites, κ_j takes values in multiples of $1/7$, and cutoffs in $[0.75, 0.85]$ select identical scaffolds; we therefore report the three distinct settings $\eta_E \in \{0.70, 0.75, 0.90\}$, corresponding to sign agreement at no fewer than five, six, and all seven training sites. As shown in Table 12, the default configuration achieves the highest mean AUC under both parcellations and the highest mean ACC under AAL. ACC under CC200 varies modestly across neighboring settings, with the best setting depending on the

Table 13: Within- and between-group gate similarity and within-group gate variation on held-out ABIDE subjects.

S_{within}	S_{between}	ΔS	ΔJ	R_{within}
$0.9649_{\pm 0.0155}$	$0.9643_{\pm 0.0160}$	$0.0005_{\pm 0.0010}$	$0.0020_{\pm 0.0028}$	$0.9697_{\pm 0.0069}$

Table 14: Top 10 scaffold edges with positive aggregate diagnostic contrasts on ABIDE under the OOD setting. Edge strength is measured by $|d_{\text{com},j}|$, the magnitude of the consensus contrast averaged over the 100 fold–run evaluations. ν_j denotes selection frequency across evaluations, and κ_j denotes source-site sign consistency within each evaluation, averaged across evaluations.

Rank	ROI 1	ROI 2	Sign	$\lvert d_{\text{com},j}\rvert$	Gate	ν_j	κ_j
1	Parietal_Inf_R	Angular_L	+	0.0757	1.000	1.00	0.92
2	Thalamus_L	Temporal_Mid_L	+	0.0715	0.976	0.99	0.89
3	Frontal_Mid_R	Angular_L	+	0.0632	0.984	1.00	0.94
4	Cingulum_Mid_L	Occipital_Sup_R	+	0.0612	0.978	0.98	0.88
5	Thalamus_L	Temporal_Sup_L	+	0.0571	0.775	0.94	0.86
6	Paracentral_Lobule_L	Pallidum_R	+	0.0526	0.979	1.00	0.89
7	Parietal_Inf_R	Angular_R	+	0.0518	0.869	0.96	0.90
8	Lingual_R	Putamen_R	+	0.0498	0.771	0.91	0.84
9	Pallidum_L	Temporal_Mid_R	+	0.0476	0.852	0.99	0.87
10	Insula_L	Paracentral_Lobule_L	+	0.0472	0.857	0.97	0.88

selection criterion. Performance declines at the most restrictive end of each tested range, indicating that overly stringent filtering can exclude informative pathways.

To complement these analyses, we examine how the learned gates vary across held-out ABIDE subjects. Held-out diagnostic labels are used only for this post-hoc analysis, and all statistics are computed within each held-out site before aggregation across LOSO folds. Let $\boldsymbol{g}_i$ denote the gate vector of subject i. We compute the mean cosine similarities between subject pairs within the same diagnostic group and across different groups, denoted by S_{within} and S_{between}, respectively, and define $\Delta S = S_{\text{within}} - S_{\text{between}}$. To compare the continuous pathway weights, we also compute weighted Jaccard similarity:

$$J_{\text{w}}(\boldsymbol{g}_i, \boldsymbol{g}_k) = \frac{\sum_p \min(g_{i,p}, g_{k,p})}{\sum_p \max(g_{i,p}, g_{k,p})}. \tag{57}$$

We analogously define ΔJ as the within-minus-between difference in mean weighted Jaccard similarity. The fraction of total gate variation occurring within diagnostic groups is measured by

$$R_{\text{within}} = \frac{\sum_{y\in\{0,1\}} \sum_{i:Y_i=y} \|\boldsymbol{g}_i - \bar{\boldsymbol{g}}_y\|_2^2}{\sum_i \|\boldsymbol{g}_i - \bar{\boldsymbol{g}}\|_2^2}, \tag{58}$$

where $\bar{\boldsymbol{g}}_y$ and $\bar{\boldsymbol{g}}$ are the group-specific and overall mean gate vectors within the held-out site. Table 13 shows high cosine and weighted Jaccard similarity both within and between diagnostic groups, with small within-minus-between differences under both measures, and most gate variation occurs within diagnostic groups rather than between their mean gate vectors. The high similarity across all subject pairs reflects the shared component introduced by the scaffold prior, which up-weights pathways with large consensus contrasts for every subject; the gating ablations in Table 10 confirm that this population weighting alone already improves over uniform gates. The near-zero group differences indicate that the remaining subject-level variation is not aligned with the diagnostic label, so the gates do not act as a shortcut classifier. Instead, the subject-conditioned term redistributes weight among scaffold pathways according to each subject's connectivity profile, while discrimination is carried out by the subsequent message passing and readout. This modulation is nonetheless beneficial, as removing the subject-conditioned term consistently lowers performance relative to the full model, and removing gating altogether leads to larger drops in the ablation study. Together, these results indicate that the population prior provides a shared pathway weighting that is stable across sites, and the subject-conditioned term supplies a modest but consistent individual adjustment on top of it.

I.4 More Interpretability Analysis

Top Positive Scaffold Edges on ABIDE. In this part, to assess biological plausibility and the stability of scaffold selection, we additionally repeat each ABIDE LOSO fold with different random seeds, whereas the main results use a single run per fold. Specifically, we summarize scaffold edges across the $E = 10$ LOSO folds, each repeated for $R = 10$ runs; for fold e and run r, let $\mathcal{S}^{(e,r)}$ denote the population scaffold estimated from the source sites. Across the resulting $ER = 100$ scaffold constructions, the selection frequency of edge j is

$$\nu_j = \frac{1}{ER} \sum_{e=1}^{E} \sum_{r=1}^{R} \mathbb{I}\Big[j \in \mathcal{S}^{(e,r)}\Big]. \tag{59}$$

For each selected edge, we first average its continuous gate over subjects from the corresponding held-out site:

$$g_j^{(e,r)} = \frac{1}{N_e} \sum_{i \in \mathcal{D}_e^{\text{test}}} g_{i,j}^{(e,r)}, \tag{60}$$

where $\mathcal{D}_e^{\text{test}}$ contains N_e test subjects and $g_{i,j}^{(e,r)}$ denotes the gate of the line-graph node corresponding to edge j. The Gate value in Table 14 averages these site-level means over the evaluations in which the edge is selected:

$$\text{Gate}_j = \frac{\sum_{(e,r):\, j \in \mathcal{S}^{(e,r)}} g_j^{(e,r)}}{\sum_{e=1}^{E} \sum_{r=1}^{R} \mathbb{I}\big[j \in \mathcal{S}^{(e,r)}\big]}. \tag{61}$$

Thus, ν_j measures recurrence across folds and runs, whereas Gate summarizes pathway weighting for held-out subjects conditional on selection. Each selected fold and run contributes equally to the reported Gate value.

Let $\kappa_j^{(e,r)}$ denote the source-site sign consistency defined in Sec. 3.1 for fold e and run r. The sign consistency reported in Table 14 is its average across all evaluations:

$$\kappa_j = \frac{1}{ER} \sum_{e=1}^{E} \sum_{r=1}^{R} \kappa_j^{(e,r)}. \tag{62}$$

We similarly average the consensus diagnostic contrasts to obtain

$$d_{\text{com},j} = \frac{1}{ER} \sum_{e=1}^{E} \sum_{r=1}^{R} d_{\text{com},j}^{(e,r)}. \tag{63}$$

An edge is termed positive when $d_{\text{com},j} > 0$. Because the diagnostic contrast is defined as ASD minus TD, positive values indicate higher residualized FC in ASD than in TD on average across evaluations. Positivity therefore refers to the direction of the label-associated contrast, not the sign of the raw Pearson correlation.

Table 14 reports the ten positive scaffold edges with the largest $|d_{\text{com},j}|$, together with their average gate values, selection frequencies, and sign consistencies. These connections involve association, sensory, and subcortical regions, including inferior parietal and angular gyri, thalamic and temporal areas, occipital cortex, insula, putamen, and pallidum. Their distribution is broadly consistent with prior ASD resting-state connectivity findings involving large-scale association networks, thalamocortical circuits, the salience network, and striatal systems Hull et al. (2017); Lee et al. (2016); Di Martino et al. (2011). Most reported edges exhibit high selection frequencies and strong sign consistency, indicating that these label-associated connections recur across LOSO folds and repeated runs.

Site-wise Scaffold Edges on ABIDE.

To examine how the selected connectivity patterns vary across LOSO folds, Fig. 7 shows, for each held-out site, the ten scaffold edges with the largest $|d_{\text{com},j}^{(e,1)}|$ in the scaffold estimated from the remaining source sites (run $r = 1$), alongside multiple anatomical views. Although the exact ROI pairs differ across folds, the selected edges repeatedly involve association, temporal, parietal, frontal, cingulate, occipital, and subcortical regions.

Table 15: Time consumption of different methods in the training stage for each epoch (in seconds).

Methods	ABIDE	REST-meta-MDD	SRPBS	ABCD
BrainOOD	0.65	2.48	0.82	0.58
BrainNetTF	0.35	1.25	0.45	0.32
AGMGC	0.32	1.18	0.42	0.30
FC-HGNN	0.15	0.35	0.18	0.12
XG-GNN	0.28	0.98	0.35	0.25
CORE	0.15	0.38	0.20	0.12

Table 16: GPU memory consumption of different methods in the training stage (in GB).

Methods	ABIDE	REST-meta-MDD	SRPBS	ABCD
BrainOOD	1.9	3.5	2.2	2.0
BrainNetTF	1.5	2.8	1.8	1.6
AGMGC	1.4	2.6	1.7	1.5
FC-HGNN	1.4	2.7	1.7	1.5
XG-GNN	1.6	3.0	1.9	1.7
CORE	1.5	3.1	1.9	1.6

I.5 Efficiency and Resource Consumption Analysis

We further evaluate the training efficiency of different methods in terms of per-epoch training time and GPU memory consumption. As shown in Tables 15 and 16, CORE achieves competitive efficiency across all datasets. In terms of training time, CORE is consistently faster than BrainOOD Xu et al. (2025), BrainNetTF Kan et al. (2022), AGMGC Noman et al. (2025), and XG-GNN Qiu et al. (2024), while remaining comparable to the lightweight FC-HGNN baseline. This efficiency stems from performing message passing on a compact scaffold-level graph, which avoids redundant computation over dense brain connectivity. For GPU memory consumption, CORE uses less memory than BrainOOD across all datasets and remains broadly comparable to XG-GNN and the lightweight baselines.

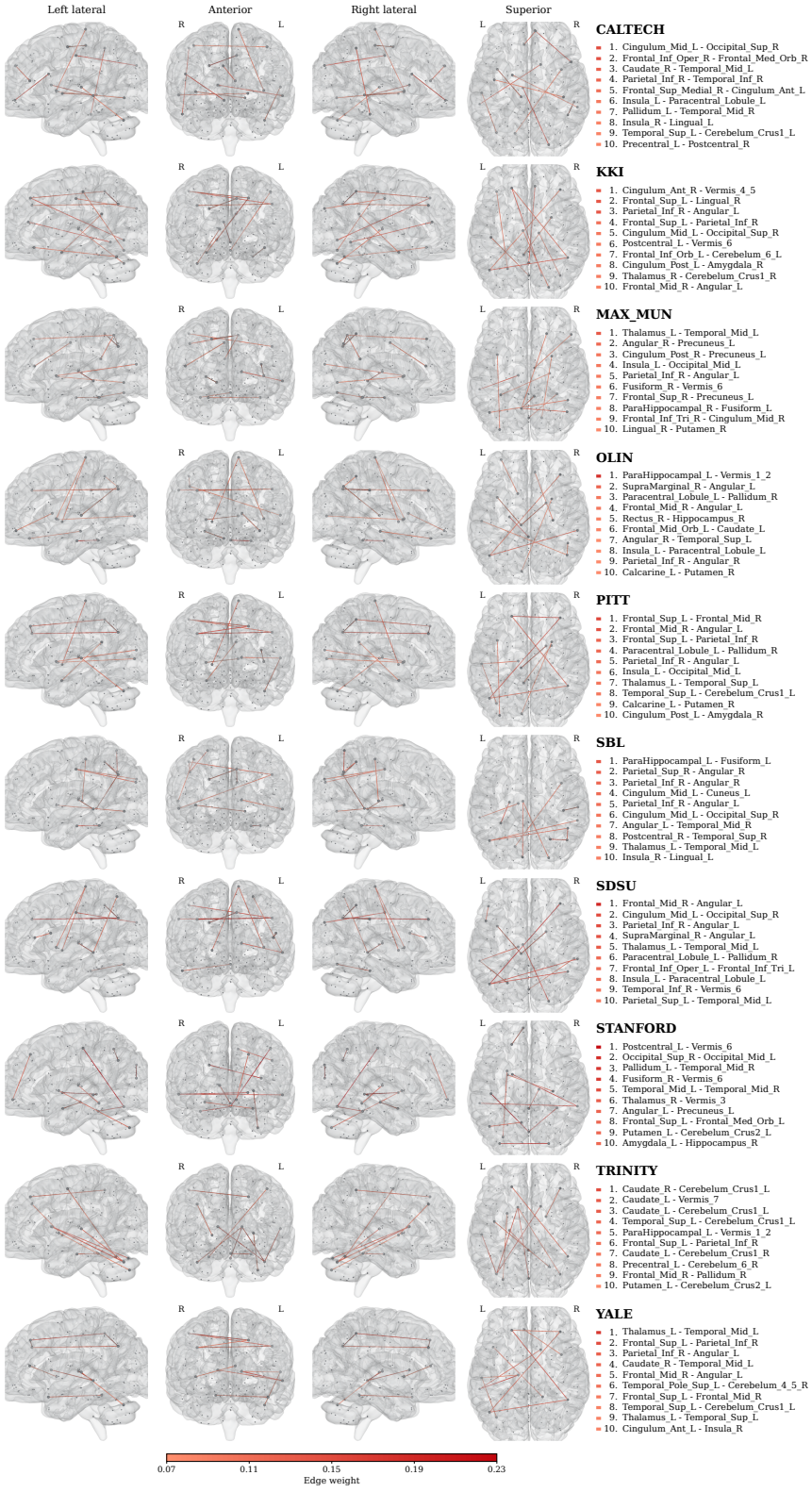


Figure 7: Scaffold edges for each held-out ABIDE site under the OOD setting. Each row lists the ten ROI pairs with the largest $|d^{(e,1)}_{\mathrm{com},j}|$ in the scaffold estimated from the remaining source sites (run $r = 1$).